GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:14:12 ; Search time 50.2677 Seconds
(without alignments)
265.240 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	423	100.0	84	21	AAB18942	Peptide derived fr
2	423	100.0	186	21	AAB18944	Peptide derived fr
3	423	100.0	540	17	AAW07871	GDU (or Grb14), a
4	386	91.3	84	21	AAB18938	Peptide derived fr
5	386	91.3	186	21	AAB18940	Peptide derived fr
6	363	85.8	174	21	AAB18943	Peptide derived fr
7	339	80.1	174	21	AAB18939	Peptide derived fr
8	212	50.1	43	21	AAB18941	Peptide derived fr
9	205	48.5	43	21	AAB18937	Peptide derived fr
10	191	45.2	80	21	AAB18954	Peptide derived fr
11	191	45.2	80	21	AAB18962	Peptide derived fr
12	191	45.2	170	21	AAB18955	Peptide derived fr
13	191	45.2	170	21	AAB18963	Peptide derived fr
14	191	45.2	182	21	AAB18956	Peptide derived fr
15	191	45.2	182	21	AAB18964	Peptide derived fr
16	191	45.2	534	16	AAR80164	Mouse signal trans
17	191	45.2	535	16	AAR86900	Human GRB-7. Homo
18	190.5	45.0	178	22	ABG02112	Novel human diagno
19	189	44.7	82	21	AAB18950	Peptide derived fr
20	189	44.7	184	21	AAB18952	Peptide derived fr
21	189	44.7	536	20	AAW83013	Human growth facto
22	189	44.7	594	22	AAB98060	Human SH2 and plec
23	189	44.7	723	22	ABG01373	Novel human diagno
24	186	44.0	82	21	AAB18946	Peptide derived fr
25	186	44.0	184	21	AAB18948	Peptide derived fr
26	186	44.0	618	16	AAR80165	Mouse signal trans
27	186	44.0	621	16	AAR85785	Human GRB-10. Hom
28	184	43.5	172	21	AAB18951	Peptide derived fr
29	184	43.5	596	22	AAB98059	Mouse Meg1/Grb10 p
30	183	43.3	172	21	AAB18947	Peptide derived fr
31	179	42.3	80	21	AAB18958	Peptide derived fr
32	179	42.3	170	21	AAB18959	Peptide derived fr
33	179	42.3	182	21	AAB18960	Peptide derived fr
34	179	42.3	498	22	AAB93348	Human protein sequ
35	179	42.3	532	23	ABG96335	Human ovarian canc
36	178	42.1	329	23	ABP41924	Human ovarian anti
37	169	40.0	43	21	AAB18949	Peptide derived fr
38	169	40.0	334	16	AAR80167	Mouse signal trans
39	169	40.0	334	16	AAR80220	GRB-7 adaptor prot
40	169	40.0	335	16	AAR80161	GRB-7 central BLM
41	167	39.5	326	16	AAR80162	GRB-10 central BLM
42	162	38.3	43	21	AAB18957	Peptide derived fr
43	161	38.1	43	21	AAB18945	Peptide derived fr
44	159	37.6	43	21	AAB18953	Peptide derived fr
45	159	37.6	43	21	AAB18961	Peptide derived fr

ALIGNMENTS

RESULT 1
AAB18942
ID AAB18942 standard; peptide; 84 AA.
XX
AC AAB18942;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 26; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 84 AA;

  Query Match   100.0%; Score 423; DB 21; Length 84;
  Best Local Similarity 100.0%; Pred. No. 6.9e-47;
  Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60

```
Qy          61 GTHGSPTASSQSSATNMAIHRSQP 84
               ||||||||||||||||||||||||
Db          61 GTHGSPTASSQSSATNMAIHRSQP 84
```

RESULT 2
AAB18944
ID AAB18944 standard; peptide; 186 AA.
XX
AC AAB18944;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 27; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 186 AA;

Query Match 100.0%; Score 423; DB 21; Length 186;

```
  Best Local Similarity   100.0%;  Pred. No. 2.2e-46;
  Matches   84;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db         61 GTHGSPTASSQSSATNMAIHRSQP 84


RESULT 3
AAW07871
ID   AAW07871 standard; Protein; 540 AA.
XX
AC   AAW07871;
XX
DT   09-FEB-1997  (first entry)
XX
DE   GDU (or Grb14), a signalling protein.
XX
KW   GDU; Grb14; signalling protein; erbB receptor; target;
KW   breast cancer; prostate cancer; tumour; PDGFr;
KW   platelet derived growth factor; receptor; wound healing;
KW   atherosclerosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          235..341
FT                   /label= PH-domain
FT                   /note= "pleckstrin-homology domain"
FT   Domain          439
FT                   /label= SH2-domain
FT                   /note= "src homology domain"
XX
PN   WO9634951-A1.
XX
PD   07-NOV-1996.
XX
PF   02-MAY-1996;   96WO-AU00258.
XX
PR   02-MAY-1995;   95AU-0002742.
XX
PA   (GARV-) GARVAN INST MEDICAL RES.
XX
PI   Daly RJ,  Sutherland RL;
XX
DR   WPI; 1996-506156/50.
DR   N-PSDB; AAT44581.
XX
PT   A new signalling protein designated GDU related to erbB receptor
PT   targets - also DNA encoding it, probes, and monoclonal antibodies
PT   for detection and treatment of breast and prostate cancer
XX
PS   Claim 3; Fig 2; 17pp; English.
```

XX
CC GDU (or Grb14) is a erB receptor target related to Grb7 and Grb10.
CC Expression of GDU is expected to serve as a prognostic indicator and
CC /or tumour marker in both breast and prostate cancer. Since
CC altered expression of GDU may also contribute to abnormal cell
CC proliferation, invasion and/or migration of cancer cells, GDU
CC singnal transduction may provide a novel therapeutic target in
CC human cancer. GDU is involved in downstream signalling initiated by
CC platelet deriv. growth factor receptor (PDGFr), and may therefore
CC provide a target in diseases or conditions in which PDGFr plays a
CC regulatory role, e.g. wound healing, fibrotic conditions and
CC atherosclerosis.
XX
SQ Sequence 540 AA;

Query Match 100.0%; Score 423; DB 17; Length 540;
Best Local Similarity 100.0%; Pred. No. 1e-45;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 414

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db        415 GTHGSPTASSQSSATNMAIHRSQP 438
```

RESULT 4
AAB18938
ID AAB18938 standard; peptide; 84 AA.
XX
AC AAB18938;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.

XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 23-24; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 84 AA;

Query Match 91.3%; Score 386; DB 21; Length 84;
Best Local Similarity 88.1%; Pred. No. 4.4e-42;
Matches 74; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | || |||||:|||||:|||||||||||||||:|||:|||||||||||||||||||||||
Db          1 QARSACSSQSVSPMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKKGCLRL 60

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              | |||||| ||||| |||:|||||
Db         61 GNHGSPTAPSQSSAVNMALHRSQP 84
```

RESULT 5
AAB18940
ID AAB18940 standard; peptide; 186 AA.
XX
AC AAB18940;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.

XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 24-25; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 186 AA;

Query Match 91.3%; Score 386; DB 21; Length 186;
Best Local Similarity 88.1%; Pred. No. 1.4e-41;
Matches 74; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | || |||||:||||||:|||||||||||||||:|||:|||||||||||||||||||||||
Db          1 QARSACSSQSVSPMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKKGCLRL 60

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              | |||||| ||||| |||:|||||
Db         61 GNHGSPTAPSQSSAVNMALHRSQP 84
```

RESULT 6
AAB18943
ID AAB18943 standard; peptide; 174 AA.
XX
AC AAB18943;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.

XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 26; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 174 AA;

  Query Match   85.8%; Score 363; DB 21; Length 174;
  Best Local Similarity   100.0%; Pred. No. 1.2e-38;
  Matches   72; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 60

Qy   73 SATNMAIHRSQP 84
              ||||||||||||
Db   61 SATNMAIHRSQP 72

RESULT 7
AAB18939
ID AAB18939 standard; peptide; 174 AA.
XX
AC AAB18939;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 24; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 174 AA;

  Query Match   80.1%; Score 339; DB 21; Length 174;
  Best Local Similarity   90.3%; Pred. No. 1.6e-35;
  Matches   65; Conservative   4; Mismatches   3; Indels   0; Gaps   0;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              ||||:|||||||||||||:|||:||||||||||||||||||||||| |||||| |||
Db          1 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKKGCLRLGNHGSPTAPSQS 60

Qy         73 SATNMAIHRSQP 84
              || |||:|||||
Db         61 SAVNMALHRSQP 72
```

RESULT 8
AAB18941
ID AAB18941 standard; peptide; 43 AA.
XX
AC AAB18941;

XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 25; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 43 AA;

Query Match 50.1%; Score 212; DB 21; Length 43;
Best Local Similarity 100.0%; Pred. No. 6.2e-20;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55
||||||||||||||||||||||||||||||||||||||||||||
Db 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

RESULT 9
AAB18937
ID AAB18937 standard; peptide; 43 AA.

XX
AC AAB18937;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 23; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 43 AA;

Query Match 48.5%; Score 205; DB 21; Length 43;
Best Local Similarity 93.0%; Pred. No. 5e-19;
Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55
              ||||:||||||||||||||:|||:||||||||||||||||||
Db          1 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 43
```

RESULT 10

AAB18954
ID AAB18954 standard; peptide; 80 AA.
XX
AC AAB18954;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 32; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 80 AA;

Query Match 45.2%; Score 191; DB 21; Length 80;
Best Local Similarity 59.7%; Pred. No. 8e-17;
Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db          13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 69
```

```
Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db          70 S-LSAAIHRTQP 80
```

RESULT 11
AAB18962
ID AAB18962 standard; peptide; 80 AA.
XX
AC AAB18962;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus muris.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 37; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 80 AA;

Query Match 45.2%; Score 191; DB 21; Length 80;
Best Local Similarity 59.7%; Pred. No. 8e-17;

```
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db          13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 69

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db          70 S-LSAAIHRTQP 80


RESULT 12
AAB18955
ID   AAB18955 standard; peptide; 170 AA.
XX
AC   AAB18955;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW   Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW   insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW   diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS   Rattus sp.
XX
PN   WO200055634-A1.
XX
PD   21-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-FR00613.
XX
PR   15-MAR-1999;   99FR-0003159.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Burnol A,  Perdereau D,  Kasus-Jacobi A,  Bereziat V,  Girard J;
XX
DR   WPI; 2000-587566/55.
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 33; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
CC   affect binding between the peptides and the insulin receptor can
CC   stimulate or inhibit tyrosine kinase activity of the receptor. The
CC   peptides are used for screening molecules for ability to treat diseases
CC   in which insulin is implicated. The peptides are used to identify agents
CC   that are potentially useful for treating insulin-associated diseases,
CC   particularly diabetes and obesity but also polycystic ovarian syndrome
```

```
CC   and syndrome X.
XX
SQ   Sequence   170 AA;

  Query Match             45.2%;  Score 191;  DB 21;  Length 170;
  Best Local Similarity   59.7%;  Pred. No. 2.4e-16;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db          1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 57

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db         58 S-LSAAIHRTQP 68


RESULT 13
AAB18963
ID   AAB18963 standard; peptide; 170 AA.
XX
AC   AAB18963;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW   Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW   insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW   diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS   Mus muris.
XX
PN   WO200055634-A1.
XX
PD   21-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-FR00613.
XX
PR   15-MAR-1999;   99FR-0003159.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Burnol A,  Perdereau D,  Kasus-Jacobi A,  Bereziat V,  Girard J;
XX
DR   WPI; 2000-587566/55.
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 37-38; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
```

CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 170 AA;

Query Match 45.2%; Score 191; DB 21; Length 170;
Best Local Similarity 59.7%; Pred. No. 2.4e-16;
Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db          1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 57

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db         58 S-LSAAIHRTQP 68
```

RESULT 14
AAB18956
ID AAB18956 standard; peptide; 182 AA.
XX
AC AAB18956;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX

PS Claim 2; Page 33-34; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 182 AA;

Query Match 45.2%; Score 191; DB 21; Length 182;
Best Local Similarity 59.7%; Pred. No. 2.6e-16;
Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db          13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 69

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db          70 S-LSAAIHRTQP 80
```

RESULT 15
AAB18964
ID AAB18964 standard; peptide; 182 AA.
XX
AC AAB18964;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus muris.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX

```
DR   WPI; 2000-587566/55.
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 38; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
CC   affect binding between the peptides and the insulin receptor can
CC   stimulate or inhibit tyrosine kinase activity of the receptor. The
CC   peptides are used for screening molecules for ability to treat diseases
CC   in which insulin is implicated. The peptides are used to identify agents
CC   that are potentially useful for treating insulin-associated diseases,
CC   particularly diabetes and obesity but also polycystic ovarian syndrome
CC   and syndrome X.
XX
SQ   Sequence   182 AA;

  Query Match             45.2%;  Score 191;  DB 21;  Length 182;
  Best Local Similarity   59.7%;  Pred. No. 2.6e-16;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:|||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 69

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db         70 S-LSAAIHRTQP 80


RESULT 16
AAR80164
ID   AAR80164 standard; peptide; 534 AA.
XX
AC   AAR80164;
XX
DT   22-APR-1996  (first entry)
XX
DE   Mouse signal transduction protein GRB-7.
XX
KW   Signal transduction protein; growth factor receptor bound; BLM domain;
KW   pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;
KW   abnormal cell development; cell movement; breast cancer; atherosclerosis.
XX
OS   Mus musculus.
XX
PN   WO9525166-A1.
XX
PD   21-SEP-1995.
XX
PF   13-MAR-1995;   95WO-US03452.
XX
```

PR 08-JUN-1994; 94US-0255785.
PR 14-MAR-1994; 94US-0212234.
XX
PA (UYNY-) UNIV NEW YORK MEDICAL CENT.
XX
PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;
XX
DR WPI; 1995-336971/43.
XX
PT Treating diseases involving abnormal signal transduction e.g. cancer
PT and psoriasis - by modulating interaction between e.g. epidermal
PT growth factor receptor and its ligand, also diagnosis and screening
PT of modulators
XX
PS Disclosure; Fig 3; 102pp; English.
XX
CC The amino acid sequence of the signal transduction protein, growth
CC factor receptor bound (GRB)-7 protein. This sequence covers from amino
CC acids 2-535 of the full length protein. The protein contains a central
CC BLM domain and within this domain a pleckstrin domain (AAR80161). The
CC central domain is flanked by a proline-rich and an SH2 domain indicating
CC that the protein is involved in signal transduction. The SH2 domain has
CC been shown to bind to the HER2 receptor protein. The protein can be used
CC to screen for cpds. which can promote or interrupt interaction of
CC proteins involved in signal transduction, esp. in neuronal diseases,
CC diseases involved with abnormal cell development and defective cell
CC movement, breast cancer, atherosclerosis, etc.
XX
SQ Sequence 534 AA;

  Query Match 45.2%; Score 191; DB 16; Length 534;
  Best Local Similarity 59.7%; Pred. No. 1.2e-15;
  Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         365 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 421

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         422 S-LSAAIHRTQP 432
```

RESULT 17
AAR86900
ID AAR86900 standard; Protein; 535 AA.
XX
AC AAR86900;
XX
DT 21-MAY-1996 (first entry)
XX
DE Human GRB-7.
XX
KW GRB-7; growth factor receptor bound; tyrosine kinase; regulation;
KW cell growth; cellular metabolism; screening; signal transduction;
KW cancer; diabetes; CORT technique; cloning of receptor targets.
XX

OS Homo sapiens.
XX
PN WO9524426-A1.
XX
PD 14-SEP-1995.
XX
PF 13-MAR-1995; 95WO-US03385.
XX
PR 11-MAR-1994; 94US-0208887.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Margolis BL, Schlessinger J, Skolnik EY;
XX
DR WPI; 1995-328235/42.
DR N-PSDB; AAT07170.
XX
PT DNA encoding tyrosine kinase-binding proteins - used to screen
PT agents capable of modulating cell growth or cellular metabolism
XX
PS Disclosure; Fig 36A-C; 215pp; English.
XX
CC Using a new cloning technique, CORT (cloning of receptor targets)
CC several new tyrosine kinase (TK) binding proteins were isolated. Growth
CC factor receptor bound proteins GRB-1,GRB-2, GRB-3, GRB-4, GRB-7 and
CC GRB-10 were isolated using this method. This sequence represents GRB-7.
CC The proteins bind to a tyrosine-phosphorylated domain of a eukaryotic
CC TK. GRB proteins can be used for screening agents which are capable
CC of modulating cell growth that occurs via signal transduction through
CC TKs. Such agents can be used to prevent or inhibit cell growth or to
CC counteract tumour development. GRB proteins are also useful for
CC identifying susceptibility to diseases asociated with alterations in
CC cellular metabolism mediated by TK pathways e.g. cancer and diabetes.
XX
SQ Sequence 535 AA;

  Query Match   45.2%; Score 191; DB 16; Length 535;
  Best Local Similarity 59.7%; Pred. No. 1.3e-15;
  Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

Qy   13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
             |:||:|:|:|||||||| |||:|| |||| |:|| ||||| || || | |
Db   366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy   73 SATNMAIHRSQP 84
             | : ||||:||
Db   423 S-LSAAIHRTQP 433

RESULT 18
ABG02112
ID ABG02112 standard; Protein; 178 AA.
XX
AC ABG02112;
XX
DT 13-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #2103.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS66299.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 32471; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 178 AA;

Query Match 45.0%; Score 190.5; DB 22; Length 178;
Best Local Similarity 78.8%; Pred. No. 3e-16;
Matches 41; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

```
Qy          7 SSQSISPM-----RSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 53
              | : : |      ||||||||||||||||||||||||||||||||||||||||
Db         79 SEEIVCPFANDGTRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 130
```

RESULT 19
AAB18950
ID AAB18950 standard; peptide; 82 AA.
XX
AC AAB18950;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 30; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 82 AA;

Query Match 44.7%; Score 189; DB 21; Length 82;

```
  Best Local Similarity   53.0%;  Pred. No. 1.5e-16;
  Matches   44;  Conservative   11;  Mismatches   26;  Indels    2;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:|||||||||||||  ||||||  ||  | |:|||  ||||:   |:
Db          1 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 59

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                 ||  :    |: : : |||:|
Db         60 NILGSQSPLHPSTLSTV-IHRTQ 81


RESULT 20
AAB18952
ID   AAB18952 standard; peptide; 184 AA.
XX
AC   AAB18952;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW   Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW   insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW   diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO200055634-A1.
XX
PD   21-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-FR00613.
XX
PR   15-MAR-1999;   99FR-0003159.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Burnol A,  Perdereau D,  Kasus-Jacobi A,  Bereziat V,  Girard J;
XX
DR   WPI; 2000-587566/55.
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 31-32; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
CC   affect binding between the peptides and the insulin receptor can
CC   stimulate or inhibit tyrosine kinase activity of the receptor. The
CC   peptides are used for screening molecules for ability to treat diseases
CC   in which insulin is implicated. The peptides are used to identify agents
CC   that are potentially useful for treating insulin-associated diseases,
```

```
CC   particularly diabetes and obesity but also polycystic ovarian syndrome
CC   and syndrome X.
XX
SQ   Sequence   184 AA;

  Query Match             44.7%;  Score 189;  DB 21;  Length 184;
  Best Local Similarity   53.0%;  Pred. No. 4.9e-16;
  Matches   44;  Conservative   11;  Mismatches   26;  Indels    2;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:||||||||||||||  |||||| || | |:||| ||||:   |:
Db          1 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 59

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                 || :    |: : : |||:|
Db         60 NILGSQSPLHPSTLSTV-IHRTQ 81


RESULT 21
AAW83013
ID   AAW83013 standard; Protein; 536 AA.
XX
AC   AAW83013;
XX
DT   29-JAN-1999  (first entry)
XX
DE   Human growth factor receptor binding insulin receptor protein.
XX
KW   Human; growth factor receptor binding insulin receptor protein;
KW   GrbIR-1; recombinant; screening.
XX
OS   Homo sapiens.
XX
PN   US5840536-A.
XX
PD   24-NOV-1998.
XX
PF   09-JUL-1997;   97US-0890094.
XX
PR   09-JUL-1996;   96US-0022703.
PR   09-JUL-1997;   97US-0890094.
XX
PA   (DUNN/) DUNNINGTON D J.
PA   (FRAN/) FRANTZ J D.
PA   (SHOE/) SHOELSON S E.
XX
PI   Dunnington DJ,  Frantz JD,  Shoelson SE;
XX
DR   WPI; 1999-034035/03.
DR   N-PSDB; AAV69865.
XX
PT   DNA encoding growth factor receptor-binding insulin receptor
PT   (GrbIR-1) polypeptide - useful in screening for compounds that
PT   modulate GrbIR-1 activity and to treat conditions related to
PT   insufficient GrbIR-1 protein function
XX
PS   Claim 4; Column 21-24; 24pp; English.
```

XX
CC The present sequence represents human growth factor receptor binding
CC insulin receptor protein (GrbIR-1). The nucleic acid encoding GrbIR-1
CC is used: (1) to produce recombinant human GrbIR-1, useful in screening
CC assays for compounds that modulate GrbIR-1 activity; and (2) to treat
CC conditions related to insufficient or altered GrbIR-1 protein function.
XX
SQ Sequence 536 AA;

Query Match 44.7%; Score 189; DB 20; Length 536;
Best Local Similarity 53.0%; Pred. No. 2.3e-15;
Matches 44; Conservative 11; Mismatches 26; Indels 2; Gaps 2;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:||||||||||||||  |||||| || | |:||| ||||:   |:
Db        353 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 411

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                 || :    |: : : |||:|
Db        412 NILGSQSPLHPSTLSTV-IHRTQ 433
```

RESULT 22
AAB98060
ID AAB98060 standard; Protein; 594 AA.
XX
AC AAB98060;
XX
DT 15-AUG-2001 (first entry)
XX
DE Human SH2 and pleckstrin homology domain-containing protein GRB10.
XX
KW Mouse; Meg1/Grb10; diabetes; transgene; transgenic animal;
KW insulin signal transduction inhibition.
XX
OS Homo sapiens.
XX
PN WO200128321-A1.
XX
PD 26-APR-2001.
XX
PF 18-AUG-2000; 2000WO-JP05546.
XX
PR 20-OCT-1999; 99JP-0298273.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Ishino F, Miyoshi N, Ishino T, Yokoyama M, Wakana S;
XX
DR WPI; 2001-300253/31.
DR N-PSDB; AAH21794.
XX
PT Transgenic non-human mammal with Meg1/Grb10 or human GRB 10 gene useful
PT as a model for onset of diabetes and for screening new diabetes
PT treatments -
XX
PS Disclosure; Page 36-38; 50pp; Japanese.

XX
CC The present invention describes a transgenic non-human mammal containing
CC the Meg1/Grb10 gene. Also described are: (1) a transgenic non human
CC mammal with human GRB10 gene; (2) a method for producing a transgenic
CC mouse; (3) method (M1) for screening for drugs for treating diabetes;
CC and (4) drugs found using (M1). The transgenic non-human mammal is
CC useful for screening for new drugs to treat diabetes. The transgenic
CC animals are models for the onset of diabetes, and may be useful in
CC discovering the mechanism for the onset of diabetes caused by inhibition
CC of insulin signal transduction, and for developing new treatments. The
CC present sequence represents the human SH2 and pleckstrin homology
CC domain-containing protein GRB10 which is given in the exemplification
CC of the present invention.
XX
SQ Sequence 594 AA;

Query Match 44.7%; Score 189; DB 22; Length 594;
Best Local Similarity 53.0%; Pred. No. 2.6e-15;
Matches 44; Conservative 11; Mismatches 26; Indels 2; Gaps 2;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:|||||||||||||  |||||| || | |:||| ||||:   |:
Db        411 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 469

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                 || :    |: : : |||:|
Db        470 NILGSQSPLHPSTLSTV-IHRTQ 491
```

RESULT 23
ABG01373
ID ABG01373 standard; Protein; 723 AA.
XX
AC ABG01373;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #1364.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
DR N-PSDB; AAS65560.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 31732; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 723 AA;

  Query Match 44.7%; Score 189; DB 22; Length 723;
  Best Local Similarity 53.0%; Pred. No. 3.5e-15;
  Matches 44; Conservative 11; Mismatches 26; Indels 2; Gaps 2;

Qy   1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
          | | | :|:||:||||||||||||| |||||| || | |:||| ||||: |:
Db   540 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 598

Qy   61 GTHGSPTASSQSSATNMAIHRSQ 83
             || : |: : : |||:|
Db   599 NILGSQSPLHPSTLSTV-IHRTQ 620

RESULT 24
AAB18946
ID AAB18946 standard; peptide; 82 AA.
XX
AC AAB18946;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus muris.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 28; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 82 AA;

  Query Match   44.0%; Score 186; DB 21; Length 82;
  Best Local Similarity   54.1%; Pred. No. 3.7e-16;
  Matches   46; Conservative   6; Mismatches   23; Indels   10; Gaps   3;

Qy   3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | | :||||:||||||||||||| |||:|| || | |:||| ||| | |:
Db   3 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 60

Qy   63 HGSPTASSQS----SATNMAIHRSQ 83
                    |||| | | |||:|
Db   61 ----ILSSQSPLHPSTLNAVIHRTQ 81

RESULT 25
AAB18948
ID AAB18948 standard; peptide; 184 AA.
XX

AC AAB18948;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus muris.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 29; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 184 AA;

Query Match 44.0%; Score 186; DB 21; Length 184;
Best Local Similarity 54.1%; Pred. No. 1.2e-15;
Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db          3 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 60

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db         61 ----ILSSQSPLHPSTLNAVIHRTQ 81
```

RESULT 26
AAR80165
ID AAR80165 standard; peptide; 618 AA.
XX
AC AAR80165;
XX
DT 22-APR-1996 (first entry)
XX
DE Mouse signal transduction protein GRB-10.
XX
KW Signal transduction protein; growth factor receptor bound; BLM domain;
KW pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;
KW abnormal cell development; cell movement; breast cancer; atherosclerosis.
XX
OS Mus musculus.
XX
PN WO9525166-A1.
XX
PD 21-SEP-1995.
XX
PF 13-MAR-1995; 95WO-US03452.
XX
PR 08-JUN-1994; 94US-0255785.
PR 14-MAR-1994; 94US-0212234.
XX
PA (UYNY-) UNIV NEW YORK MEDICAL CENT.
XX
PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;
XX
DR WPI; 1995-336971/43.
XX
PT Treating diseases involving abnormal signal transduction e.g. cancer
PT and psoriasis - by modulating interaction between e.g. epidermal
PT growth factor receptor and its ligand, also diagnosis and screening
PT of modulators
XX
PS Disclosure; Fig 3; 102pp; English.
XX
CC The amino acid sequence of the signal transduction protein, growth
CC factor receptor bound (GRB)-10 protein. This sequence covers from amino
CC acids 4-621 of the full length protein. The protein contains a central
CC BLM domain and within this domain a pleckstrin domain (AAR80162). The
CC central domain is flanked by a proline-rich and an SH2 domain indicating
CC that the protein is involved in signal transduction. The SH2 domain has
CC been shown to bind to the HER2 receptor protein. The protein can be used
CC to screen for cpds. which can promote or interrupt interaction of
CC proteins involved in signal transduction, esp. in neuronal diseases,
CC diseases involved with abnormal cell development and defective cell
CC movement, breast cancer, atherosclerosis, etc.
XX
SQ Sequence 618 AA;

  Query Match 44.0%; Score 186; DB 16; Length 618;
  Best Local Similarity 54.1%; Pred. No. 6.9e-15;
  Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        437 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 494

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db        495 ----ILSSQSPLHPSTLNAVIHRTQ 515
```

RESULT 27
AAR85785
ID   AAR85785 standard; Protein; 621 AA.
XX
AC   AAR85785;
XX
DT   16-MAY-1996 (first entry)
XX
DE   Human GRB-10.
XX
KW   GRB-10; growth factor receptor bound; tyrosine kinase; regulation;
KW   cell growth; cellular metabolism; screening; signal transduction;
KW   cancer; diabetes; CORT technique; cloning of receptor targets.
XX
OS   Homo sapiens.
XX
PN   WO9524426-A1.
XX
PD   14-SEP-1995.
XX
PF   13-MAR-1995;   95WO-US03385.
XX
PR   11-MAR-1994;   94US-0208887.
XX
PA   (UYNY) UNIV NEW YORK STATE.
XX
PI   Margolis BL, Schlessinger J, Skolnik EY;
XX
DR   WPI; 1995-328235/42.
DR   N-PSDB; AAT03197.
XX
PT   DNA encoding tyrosine kinase-binding proteins - used to screen
PT   agents capable of modulating cell growth or cellular metabolism
XX
PS   Claim 1; Fig 38; 215pp; English.
XX
CC   Using a new cloning technique, CORT (cloning of receptor targets)
CC   several new tyrosine kinase (TK) binding proteins were isolated. Growth
CC   factor receptor bound proteins GRB-1,GRB-2, GRB-3, GRB-4, GRB-7 and
CC   GRB-10 were isolated using this method. This sequence represents GRB-10.
CC   The proteins bind to a tyrosine-phosphorylated domain of a eukaryotic
CC   TK. GRB proteins can be used for screening agents which are capable
CC   of modulating cell growth that occurs via signal transduction through
CC   TKs. Such agents can be used to prevent or inhibit cell growth or to
CC   counteract tumour development. GRB proteins are also useful for
CC   identifying susceptibility to diseases asociated with alterations in
CC   cellular metabolism mediated by TK pathways e.g. cancer and diabetes.

```
XX
SQ   Sequence   621 AA;

  Query Match             44.0%;  Score 186;  DB 16;  Length 621;
  Best Local Similarity   54.1%;  Pred. No. 6.9e-15;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                     ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518


RESULT 28
AAB18951
ID   AAB18951 standard; peptide; 172 AA.
XX
AC   AAB18951;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW   Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW   insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW   diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO200055634-A1.
XX
PD   21-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-FR00613.
XX
PR   15-MAR-1999;   99FR-0003159.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Burnol A,  Perdereau D,  Kasus-Jacobi A,  Bereziat V,  Girard J;
XX
DR   WPI; 2000-587566/55.
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 30-31; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
CC   affect binding between the peptides and the insulin receptor can
```

CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 172 AA;

Query Match 43.5%; Score 184; DB 21; Length 172;
Best Local Similarity 57.7%; Pred. No. 2e-15;
Matches 41; Conservative 10; Mismatches 18; Indels 2; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|||||||||||||  |||||| || | |:||| ||||:   |:   || :    |
Db          1 PVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRMNILGSQSPLHPS 59

Qy         73 SATNMAIHRSQ 83
              : : : |||:|
Db         60 TLSTV-IHRTQ 69
```

RESULT 29
AAB98059
ID AAB98059 standard; Protein; 596 AA.
XX
AC AAB98059;
XX
DT 15-AUG-2001 (first entry)
XX
DE Mouse Meg1/Grb10 protein sequence SEQ ID NO:2.
XX
KW Mouse; Meg1/Grb10; diabetes; transgene; transgenic animal;
KW insulin signal transduction inhibition.
XX
OS Mus sp.
XX
PN WO200128321-A1.
XX
PD 26-APR-2001.
XX
PF 18-AUG-2000; 2000WO-JP05546.
XX
PR 20-OCT-1999; 99JP-0298273.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Ishino F, Miyoshi N, Ishino T, Yokoyama M, Wakana S;
XX
DR WPI; 2001-300253/31.
DR N-PSDB; AAH21792, AAH21793.
XX
PT Transgenic non-human mammal with Meg1/Grb10 or human GRB 10 gene useful
PT as a model for onset of diabetes and for screening new diabetes
PT treatments -
XX
PS Claim 2; Page 30-31; 50pp; Japanese.

XX
CC The present invention describes a transgenic non-human mammal containing
CC the Meg1/Grb10 gene. Also described are: (1) a transgenic non human
CC mammal with human GRB10 gene; (2) a method for producing a transgenic
CC mouse; (3) method (M1) for screening for drugs for treating diabetes;
CC and (4) drugs found using (M1). The transgenic non-human mammal is
CC useful for screening for new drugs to treat diabetes. The transgenic
CC animals are models for the onset of diabetes, and may be useful in
CC discovering the mechanism for the onset of diabetes caused by inhibition
CC of insulin signal transduction, and for developing new treatments. The
CC present sequence represents a specifically claimed mouse Meg1/Grb10
CC protein sequence from the present invention.
XX
SQ Sequence 596 AA;

Query Match 43.5%; Score 184; DB 22; Length 596;
Best Local Similarity 54.1%; Pred. No. 1.2e-14;
Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        415 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGRTRMN- 472

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                     ||||    |  |  |||:|
Db        473 ----ILSSQSPLHPSTLNAVIHRTQ 493
```

RESULT 30
AAB18947
ID AAB18947 standard; peptide; 172 AA.
XX
AC AAB18947;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus muris.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.

```
XX
PT   Fragments of Grb family proteins to identify compounds are useful in
PT   treating insulin-associated diseases, particularly diabetes and obesity
PT    -
XX
PS   Claim 2; Page 28-29; 46pp; French.
XX
CC   B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC   region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC   PIR is the actual binding region but its effect is about 10 times
CC   greater in presence of SH2 (which by itself is inactive). Agents that
CC   affect binding between the peptides and the insulin receptor can
CC   stimulate or inhibit tyrosine kinase activity of the receptor. The
CC   peptides are used for screening molecules for ability to treat diseases
CC   in which insulin is implicated. The peptides are used to identify agents
CC   that are potentially useful for treating insulin-associated diseases,
CC   particularly diabetes and obesity but also polycystic ovarian syndrome
CC   and syndrome X.
XX
SQ   Sequence   172 AA;

  Query Match             43.3%;  Score 183;  DB 21;  Length 172;
  Best Local Similarity   58.7%;  Pred. No. 2.7e-15;
  Matches   44;  Conservative    5;  Mismatches   16;  Indels   10;  Gaps    3;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              ||||:|||||||||||||  |||:|| || | |:||| |||  |  |:        ||||
Db          1 PMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN-----ILSSQS 54

Qy         73 ----SATNMAIHRSQ 83
                   |  |  |||:|
Db         55 PLHPSTLNAVIHRTQ 69


RESULT 31
AAB18958
ID   AAB18958 standard; peptide; 80 AA.
XX
AC   AAB18958;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW   Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW   insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW   diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO200055634-A1.
XX
PD   21-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-FR00613.
XX
PR   15-MAR-1999;   99FR-0003159.
```

XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 34-35; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 80 AA;

Query Match 42.3%; Score 179; DB 21; Length 80;
Best Local Similarity 59.2%; Pred. No. 2.9e-15;
Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| |||||||:||  |||||   ||     |  :| :
Db         13 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 69

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db         70 S-LSAAIHRTQ 79
```

RESULT 32
AAB18959
ID AAB18959 standard; peptide; 170 AA.
XX
AC AAB18959;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.

XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 35; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 170 AA;

Query Match 42.3%; Score 179; DB 21; Length 170;
Best Local Similarity 59.2%; Pred. No. 8.6e-15;
Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| ||||||:||  |||||   ||     |  :| :
Db          1 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 57

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db         58 S-LSAAIHRTQ 67
```

RESULT 33
AAB18960
ID AAB18960 standard; peptide; 182 AA.
XX
AC AAB18960;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 35-36; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 182 AA;

  Query Match   42.3%; Score 179; DB 21; Length 182;
  Best Local Similarity   59.2%; Pred. No. 9.5e-15;
  Matches   42; Conservative   8; Mismatches   17; Indels   4; Gaps   2;

Qy   13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:|||||||| |||||| ||||||:|| ||||| || | :| :
Db   13 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 69

Qy   73 SATNMAIHRSQ 83
              | : ||||:|
Db   70 S-LSAAIHRTQ 79

RESULT 34
AAB93348
ID AAB93348 standard; Protein; 498 AA.
XX
AC AAB93348;

XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12468.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12468; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 498 AA;

Query Match 42.3%; Score 179; DB 22; Length 498;
Best Local Similarity 59.2%; Pred. No. 4.1e-14;
Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| ||||||:||  |||||   ||     |  :| :
Db        329 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 385

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db        386 S-LSAAIHRTQ 395
```

RESULT 35
ABG96335
ID ABG96335 standard; Protein; 532 AA.
XX
AC ABG96335;
XX
DT 11-DEC-2002 (first entry)
XX
DE Human ovarian cancer marker M447.
XX
KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW central nervous system disorder; bacterial meningitis; viral meningitis;
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW brain herniation; inflammation; encephalitis; testicular disorder;
KW nontuberculous granulomatous orchitis; connective tissue disorder;
KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW histological type; carcinogenic; ovarian cancer marker.
XX
OS Homo sapiens.
XX
PN WO200271928-A2.
XX
PD 19-SEP-2002.
XX
PF 14-MAR-2002; 2002WO-US07826.
XX
PR 14-MAR-2001; 2001US-276025P.
PR 14-MAR-2001; 2001US-276026P.
PR 10-AUG-2001; 2001US-311732P.
PR 19-SEP-2001; 2001US-323580P.
PR 26-SEP-2001; 2001US-324967P.
PR 26-SEP-2001; 2001US-325102P.
PR 26-SEP-2001; 2001US-325149P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
DR WPI; 2002-723277/78.

DR N-PSDB; ABS76431.
XX
PT Assessing whether a patient is afflicted with ovarian cancer, useful in
PT assessing the stage or progression of the disease, comprises comparing
PT the expression level of a cancer marker in a sample from a patient and
PT from a non cancer patient -
XX
PS Disclosure; Page 245-246; 481pp; English.
XX
CC The present invention relates to a new method for assessing whether a
CC patient is afflicted with ovarian cancer. The method involves comparing
CC the expression level of a marker in a patient sample and the normal level
CC of expression of the marker in a control non-ovarian cancer sample, where
CC the marker is selected from 363 cancer markers described in the
CC specification. The method of the invention is useful in diagnosing or
CC characterising cancer, in detecting the presence of cancer as early as
CC possible, and the recurrence of ovarian cancer. The method may also be of
CC particular use with patients having an enhanced risk of developing
CC ovarian cancer (e.g. patients having a familial history of ovarian
CC cancer). The cancer markers may be used in the management and treatment
CC of e.g. brain and central nervous system disorders (e.g. bacterial and
CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
CC inflammations (e.g. bacterial or viral meningitis or encephalitis),
CC testicular disorders (e.g. nontuberculous granulomatous orchitis),
CC connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC disease or atherosclerosis). The compositions and methods may also be
CC used in assessing the histological type of neoplasm associated with
CC ovarian cancer, monitoring the progression of ovarian cancer,
CC determining whether ovarian cancer has metastasized or is likely to
CC metastasize, selecting a composition for inhibiting ovarian cancer,
CC assessing the ovarian carcinogenic potential of a compound, or
CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The
CC present amino acid sequence represents one of the ovarian cancer markers
CC described in the invention.
XX
SQ Sequence 532 AA;

  Query Match   42.3%; Score 179; DB 23; Length 532;
  Best Local Similarity   59.2%; Pred. No. 4.5e-14;
  Matches   42; Conservative   8; Mismatches   17; Indels   4; Gaps   2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| ||||||:||  |||||   ||     |  :| :
Db        363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 419

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db        420 S-LSAAIHRTQ 429
```

RESULT 36
ABP41924
ID ABP41924 standard; Protein; 329 AA.
XX
AC ABP41924;
XX

DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HODKM52, SEQ ID NO:3056.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
DR N-PSDB; ABQ55001.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11; SEQ ID No 3056; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders

CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 329 AA;

Query Match 42.1%; Score 178; DB 23; Length 329;
Best Local Similarity 59.2%; Pred. No. 3e-14;
Matches 42; Conservative 7; Mismatches 18; Indels 4; Gaps 2;

```
Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| |||||||:||  |||||   ||     |  :|
Db        160 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGX 216

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db        217 S-LSAAIHRTQ 226
```

RESULT 37
AAB18949
ID AAB18949 standard; peptide; 43 AA.
XX
AC AAB18949;
XX
DT 08-FEB-2001 (first entry)
XX
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Homo sapiens.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR00613.
XX
PR 15-MAR-1999; 99FR-0003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX

PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity
PT -
XX
PS Claim 2; Page 30; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.
CC PIR is the actual binding region but its effect is about 10 times
CC greater in presence of SH2 (which by itself is inactive). Agents that
CC affect binding between the peptides and the insulin receptor can
CC stimulate or inhibit tyrosine kinase activity of the receptor. The
CC peptides are used for screening molecules for ability to treat diseases
CC in which insulin is implicated. The peptides are used to identify agents
CC that are potentially useful for treating insulin-associated diseases,
CC particularly diabetes and obesity but also polycystic ovarian syndrome
CC and syndrome X.
XX
SQ Sequence 43 AA;

  Query Match 40.0%; Score 169; DB 21; Length 43;
  Best Local Similarity 76.7%; Pred. No. 2.3e-14;
  Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55
       |:||:|||||||||||||| |||||| || | |:||| ||||:
Db   1 PVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKR 43

RESULT 38
AAR80167
ID AAR80167 standard; peptide; 334 AA.
XX
AC AAR80167;
XX
DT 22-APR-1996 (first entry)
XX
DE Mouse signal transduction protein GRB-7 residues 95-428.
XX
KW Signal transduction protein; growth factor receptor bound; BLM domain;
KW pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;
KW abnormal cell development; cell movement; breast cancer; atherosclerosis.
XX
OS Mus musculus.
XX
PN WO9525166-A1.
XX
PD 21-SEP-1995.
XX
PF 13-MAR-1995; 95WO-US03452.
XX
PR 08-JUN-1994; 94US-0255785.
PR 14-MAR-1994; 94US-0212234.
XX
PA (UYNY-) UNIV NEW YORK MEDICAL CENT.
XX
PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;

```
XX
DR   WPI; 1995-336971/43.
XX
PT   Treating diseases involving abnormal signal transduction e.g. cancer
PT   and psoriasis - by modulating interaction between e.g. epidermal
PT   growth factor receptor and its ligand, also diagnosis and screening
PT   of modulators
XX
PS   Claim 15; Fig 3; 102pp; English.
XX
CC   The amino acid sequence of the signal transduction protein, growth
CC   factor receptor bound (GRB)-7 protein. This sequence covers from amino
CC   acids 95-428 of the full length protein.  The protein contains a central
CC   BLM domain and within this domain a pleckstrin domain (AAR80161).  The
CC   central domain is flanked by a proline-rich and an SH2 domain indicating
CC   that the protein is involved in signal transduction. The SH2 domain has
CC   been shown to bind to the HER2 receptor protein. The protein can be used
CC   to screen for cpds. which can promote or interrupt interaction of
CC   proteins involved in signal transduction, esp. in neuronal diseases,
CC   diseases involved with abnormal cell development and defective cell
CC   movement, breast cancer, atherosclerosis, etc.
XX
SQ   Sequence   334 AA;

  Query Match             40.0%;  Score 169;  DB 16;  Length 334;
  Best Local Similarity   58.1%;  Pred. No. 4.5e-13;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db        272 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 331

Qy         73 SA 74
              :|
Db        332 AA 333


RESULT 39
AAR80220
ID   AAR80220 standard; peptide; 334 AA.
XX
AC   AAR80220;
XX
DT   29-APR-1996  (first entry)
XX
DE   GRB-7 adaptor protein.
XX
KW   PTK; oncogene; identification; detection; breast cancer; receptor;
KW   complex; adaptor; HER-2; GRB.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "unspecified amino acid"
FT   Misc-difference 4
FT                   /note= "unspecified amino acid"
```

```
FT   Misc-difference 5
FT                   /note= "unspecified amino acid"
XX
PN   WO9524205-A1.
XX
PD   14-SEP-1995.
XX
PF   07-MAR-1995;   95WO-US02787.
XX
PR   07-MAR-1994;   94US-0207575.
XX
PA   (UYNY-) UNIV NEW YORK MEDICAL CENT.
XX
PI   Margolis BL;
XX
DR   WPI; 1995-328097/42.
XX
PT   Identification of cpds. for modulating an oncogenic disorder esp.
PT   breast cancer  - by exposing potential agents to a receptor protein
PT   tyrosine kinase polypeptide/adaptor polypeptide complex
XX
PS   Disclosure; Fig 8B; 112pp; English.
XX
CC   Conserved motifs of the protein tyrosine kinase (PTK) catalytic
CC   domain may be complexed with an adaptor polypeptide  to give a
CC   receptor protein tyrosine kinase/adaptor protein (RpTKp/Ap) complex.
CC   The adaptor protein is a member of the SH2 and SH3 contg. family of
CC   adaptor proteins and is pref. a GRB-7 adaptor protein. A preferred
CC   compound of the invention is an HER2/GRB-7 complex. The complexes
CC   can be used to screen for candidate compounds for modulating
CC   oncogenic disorders in partic. breast cancer.
XX
SQ   Sequence   334 AA;

  Query Match             40.0%;  Score 169;  DB 16;  Length 334;
  Best Local Similarity   58.1%;  Pred. No. 4.5e-13;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db        272 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 331

Qy         73 SA 74
              :|
Db        332 AA 333


RESULT 40
AAR80161
ID   AAR80161 standard; peptide; 335 AA.
XX
AC   AAR80161;
XX
DT   22-APR-1996  (first entry)
XX
DE   GRB-7 central BLM domain.
XX
```

KW Signal transduction protein; growth factor receptor bound; BLM domain;
KW pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;
KW abnormal cell development; cell movement; breast cancer; atherosclerosis.
XX
OS Mus musculus.
XX
PN WO9525166-A1.
XX
PD 21-SEP-1995.
XX
PF 13-MAR-1995; 95WO-US03452.
XX
PR 08-JUN-1994; 94US-0255785.
PR 14-MAR-1994; 94US-0212234.
XX
PA (UYNY-) UNIV NEW YORK MEDICAL CENT.
XX
PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;
XX
DR WPI; 1995-336971/43.
XX
PT Treating diseases involving abnormal signal transduction e.g. cancer
PT and psoriasis - by modulating interaction between e.g. epidermal
PT growth factor receptor and its ligand, also diagnosis and screening
PT of modulators
XX
PS Disclosure; Fig 2; 102pp; English.
XX
CC The amino acid sequence of the central domain of the signal transduction
CC protein, growth factor receptor bound (GRB)-7 protein. The protein
CC contains a central BLM domain and within this domain a pleckstrin domain.
CC The central domain is flanked by a proline-rich and an SH2 domain
CC indicating that the protein is involved in signal transduction. The SH2
CC domain has been shown to bind to the HER2 receptor protein. The protein
CC can be used to screen for cpds. which can promote or interrupt
CC interaction of proteins involved in signal transduction, esp. in neuronal
CC diseases, diseases involved with abnormal cell development and defective
CC cell movement, breast cancer, atherosclerosis, etc.
XX
SQ Sequence 335 AA;

Query Match 40.0%; Score 169; DB 16; Length 335;
Best Local Similarity 58.1%; Pred. No. 4.5e-13;
Matches 36; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db         273 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 332

Qy          73 SA 74
               :|
Db         333 AA 334
```

Search completed: January 13, 2004, 16:20:52
Job time : 50.2677 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:18:37 ; Search time 19.8425 Seconds
(without alignments)
179.116 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	423	100.0	540	4	US-08-945-771-2	Sequence 2, Appli
2	191	45.2	534	3	US-08-866-381A-5	Sequence 5, Appli
3	191	45.2	535	1	US-07-906-349A-10	Sequence 10, Appl
4	191	45.2	535	1	US-08-167-035-10	Sequence 10, Appl
5	191	45.2	535	1	US-08-208-887A-10	Sequence 10, Appl
6	191	45.2	535	2	US-08-539-005-10	Sequence 10, Appl
7	191	45.2	535	4	US-09-280-598-10	Sequence 10, Appl
8	191	45.2	535	4	US-08-945-771-3	Sequence 3, Appli
9	189	44.7	536	2	US-08-890-094-2	Sequence 2, Appli
10	189	44.7	548	2	US-08-890-094-18	Sequence 18, Appl
11	186	44.0	618	3	US-08-866-381A-6	Sequence 6, Appli

12	186	44.0	621	1	US-08-208-887A-49	Sequence 49, Appl
13	186	44.0	621	4	US-09-280-598-18	Sequence 18, Appl
14	186	44.0	621	4	US-08-945-771-4	Sequence 4, Appli
15	169	40.0	334	3	US-08-472-595-9	Sequence 9, Appli
16	169	40.0	334	3	US-08-207-575A-9	Sequence 9, Appli
17	169	40.0	335	3	US-08-866-381A-1	Sequence 1, Appli
18	169	40.0	335	4	US-09-280-598-51	Sequence 51, Appl
19	167	39.5	326	3	US-08-866-381A-2	Sequence 2, Appli
20	167	39.5	326	4	US-09-280-598-52	Sequence 52, Appl
21	72	17.0	1151	4	US-09-023-905A-4	Sequence 4, Appli
22	69	16.3	243	4	US-09-252-991A-28884	Sequence 28884, A
23	66	15.6	863	4	US-09-252-991A-19574	Sequence 19574, A
24	65	15.4	653	4	US-09-198-452A-439	Sequence 439, App
25	61.5	14.5	388	1	US-08-429-742-4	Sequence 4, Appli
26	61	14.4	374	3	US-08-821-994-68	Sequence 68, Appl
27	61	14.4	384	4	US-09-252-991A-21729	Sequence 21729, A
28	60.5	14.3	2860	2	US-08-826-267-2	Sequence 2, Appli
29	60	14.2	169	4	US-09-252-991A-20992	Sequence 20992, A
30	60	14.2	169	4	US-09-252-991A-22999	Sequence 22999, A
31	60	14.2	169	4	US-09-252-991A-25204	Sequence 25204, A
32	60	14.2	169	4	US-09-252-991A-26569	Sequence 26569, A
33	60	14.2	169	4	US-09-252-991A-31908	Sequence 31908, A
34	60	14.2	310	4	US-09-598-747-27	Sequence 27, Appl
35	60	14.2	480	4	US-09-107-532A-6160	Sequence 6160, Ap
36	60	14.2	950	4	US-09-328-352-4668	Sequence 4668, Ap
37	59.5	14.1	1297	4	US-09-107-532A-4552	Sequence 4552, Ap
38	59.5	14.1	1346	3	US-09-320-878-4	Sequence 4, Appli
39	59.5	14.1	1346	4	US-09-141-908-5	Sequence 5, Appli
40	59.5	14.1	1346	4	US-09-657-440-4	Sequence 4, Appli
41	59	13.9	638	4	US-09-252-991A-24325	Sequence 24325, A
42	59	13.9	1024	4	US-09-562-737-85	Sequence 85, Appl
43	58.5	13.8	382	3	US-09-586-719-8	Sequence 8, Appli
44	58.5	13.8	521	3	US-08-956-322-4	Sequence 4, Appli
45	58.5	13.8	652	3	US-09-110-116-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-945-771-2
; Sequence 2, Application US/08945771
; Patent No. 6465623
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger J
; APPLICANT: Sutherland, Robert L
; TITLE OF INVENTION: GDU, A novel signalling protein
; FILE REFERENCE: 273402001700
; CURRENT APPLICATION NUMBER: US/08/945,771
; CURRENT FILING DATE: 1998-04-22
; EARLIER APPLICATION NUMBER: PCT/US96/00258
; EARLIER FILING DATE: 1996-MAY-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 540
;   TYPE: PRT

; ORGANISM: Homo sapiens
US-08-945-771-2

  Query Match   100.0%; Score 423; DB 4; Length 540;
  Best Local Similarity   100.0%; Pred. No. 3e-48;
  Matches   84; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 414

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db        415 GTHGSPTASSQSSATNMAIHRSQP 438
```

RESULT 2
US-08-866-381A-5
; Sequence 5, Application US/08866381A
; Patent No. 6045797
; GENERAL INFORMATION:
;   APPLICANT: Ben Lewis Margolis
;   APPLICANT: Joseph Schlessinger
;   TITLE OF INVENTION: METHODS FOR TREATMENT OR DIAGNOSIS
;   TITLE OF INVENTION: OF DISEASES OR CONDITIONS ASSOCIATED
;   TITLE OF INVENTION: WITH A BLM DOMAIN
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Lyon & Lyon
;   STREET: 633 West Fifth Street
;   STREET: Suite 4700
;   CITY: Los Angeles
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 90071-2066
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   MEDIUM TYPE: storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: IBM P.C. DOS 5.0
;   SOFTWARE: FastSEQ for Windows 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/866,381A
;   FILING DATE: May 30, 1997
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/212,234
;   FILING DATE: March 14, 1994
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warburg, Richard J.
;   REGISTRATION NUMBER: 32,327
;   REFERENCE/DOCKET NUMBER: 226/043
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (213) 489-1600
;   TELEFAX: (213) 955-0440

```
;       TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  534 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;       OTHER INFORMATION:  GRB-7
US-08-866-381A-5

  Query Match             45.2%;  Score 191;  DB 3;  Length 534;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         365 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 421

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         422 S-LSAAIHRTQP 432


RESULT 3
US-07-906-349A-10
; Sequence 10, Application US/07906349A
; Patent No. 5434064
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnik, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
KINASES AND
;    TITLE OF INVENTION:  TARGET PROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Browdy and Neimark
;       STREET:  419 Seventh Street, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20004
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/906,349A
;       FILING DATE:  30-JUN-1992
;       CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/643,237
;       FILING DATE:  18-JAN-1991
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  535 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-906-349A-10

  Query Match             45.2%;  Score 191;  DB 1;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 4
US-08-167-035-10
; Sequence 10, Application US/08167035
; Patent No. 5618691
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnick, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  10036-2711
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/167,035
;      FILING DATE:  16-DEC-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-062
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  535 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-167-035-10

  Query Match             45.2%;  Score 191;  DB 1;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:|||||||||   |||:|| |||| |:||  |||||    ||      ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 5
US-08-208-887A-10
; Sequence 10, Application US/08208887A
; Patent No. 5677421
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnick, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  10036-2711
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/208,887A
;      FILING DATE:  11-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-063
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  535 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-208-887A-10

  Query Match             45.2%;  Score 191;  DB 1;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:|||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 6
US-08-539-005-10
; Sequence 10, Application US/08539005
; Patent No. 5858686
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnick, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  10036-2711
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/539,005
;      FILING DATE:  4-OCT-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/167,035
;      FILING DATE:  16-DEC-1993
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  535 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-539-005-10

  Query Match             45.2%;  Score 191;  DB 2;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 7
US-09-280-598-10
; Sequence 10, Application US/09280598
; Patent No. 6391584
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnik, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  App, Harold
;    TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/280,598
;      FILING DATE:
```

```
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/252,820
;       FILING DATE:  02-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7683-067
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  535 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-09-280-598-10

  Query Match             45.2%;  Score 191;  DB 4;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 8
US-08-945-771-3
; Sequence 3, Application US/08945771
; Patent No. 6465623
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger J
;  APPLICANT: Sutherland, Robert L
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001700
;  CURRENT APPLICATION NUMBER: US/08/945,771
;  CURRENT FILING DATE: 1998-04-22
;  EARLIER APPLICATION NUMBER: PCT/US96/00258
;  EARLIER FILING DATE: 1996-MAY-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 535
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-08-945-771-3

  Query Match             45.2%;  Score 191;  DB 4;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 4.9e-17;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;
```

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         423 S-LSAAIHRTQP 433


RESULT 9
US-08-890-094-2
; Sequence 2, Application US/08890094
; Patent No. 5840536
;  GENERAL INFORMATION:
;    APPLICANT:  SmithKline Beecham Corporation and Harvard University
;    TITLE OF INVENTION:  GROWTH FACTOR RECEPTOR-BINDING INSULIN RECEPTOR
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,094
;      FILING DATE:  09-JULY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/022,703
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baumeister, Kirk
;      REGISTRATION NUMBER:  33,833
;      REFERENCE/DOCKET NUMBER:  P50508P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5096
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
US-08-890-094-2
```

```
  Query Match             44.7%;  Score 189;  DB 2;  Length 536;
  Best Local Similarity   53.0%;  Pred. No. 9.1e-17;
  Matches   44;  Conservative   11;  Mismatches   26;  Indels    2;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:|||||||||||||  ||||||  ||  | |:|||  ||||:   |:
Db        353 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 411

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                    ||  :    |: : : |||:|
Db        412 NILGSQSPLHPSTLSTV-IHRTQ 433


RESULT 10
US-08-890-094-18
; Sequence 18, Application US/08890094
; Patent No. 5840536
;  GENERAL INFORMATION:
;    APPLICANT:  SmithKline Beecham Corporation and Harvard University
;    TITLE OF INVENTION:  GROWTH FACTOR RECEPTOR-BINDING INSULIN RECEPTOR
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,094
;      FILING DATE:  09-JULY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/022,703
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baumeister, Kirk
;      REGISTRATION NUMBER:  33,833
;      REFERENCE/DOCKET NUMBER:  P50508P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5096
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
```

```
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  N-terminal
;     ORIGINAL SOURCE:
US-08-890-094-18

  Query Match             44.7%;  Score 189;  DB 2;  Length 548;
  Best Local Similarity   53.0%;  Pred. No. 9.4e-17;
  Matches   44;  Conservative   11;  Mismatches   26;  Indels    2;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | |    |   :|:||:||||||||||||||  |||||| || | |:||| ||||:   |:
Db        365 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 423

Qy         61 GTHGSPTASSQSSATNMAIHRSQ 83
                 || :    |: : : |||:|
Db        424 NILGSQSPLHPSTLSTV-IHRTQ 445


RESULT 11
US-08-866-381A-6
; Sequence 6, Application US/08866381A
; Patent No. 6045797
;  GENERAL INFORMATION:
;    APPLICANT:  Ben Lewis Margolis
;    APPLICANT:  Joseph Schlessinger
;    TITLE OF INVENTION:  METHODS FOR TREATMENT OR DIAGNOSIS
;    TITLE OF INVENTION:  OF DISEASES OR CONDITIONS ASSOCIATED
;    TITLE OF INVENTION:  WITH A BLM DOMAIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,381A
;      FILING DATE:  May 30, 1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/212,234
;      FILING DATE:  March 14, 1994
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  226/043
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  618 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FEATURE:
;       OTHER INFORMATION:  GRB-10
US-08-866-381A-6

  Query Match             44.0%;  Score 186;  DB 3;  Length 618;
  Best Local Similarity   54.1%;  Pred. No. 2.8e-16;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        437 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 494

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                      ||||    |  |  |||:|
Db        495 ----ILSSQSPLHPSTLNAVIHRTQ 515


RESULT 12
US-08-208-887A-49
; Sequence 49, Application US/08208887A
; Patent No. 5677421
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnick, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  10036-2711
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/208,887A
;      FILING DATE:  11-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
```

```
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7683-063
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  49:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  621 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-208-887A-49

  Query Match             44.0%;  Score 186;  DB 1;  Length 621;
  Best Local Similarity   54.1%;  Pred. No. 2.9e-16;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518


RESULT 13
US-09-280-598-18
; Sequence 18, Application US/09280598
; Patent No. 6391584
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnik, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  App, Harold
;    TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/280,598
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/252,820
;      FILING DATE:  02-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-067
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  621 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-09-280-598-18

  Query Match             44.0%;  Score 186;  DB 4;  Length 621;
  Best Local Similarity   54.1%;  Pred. No. 2.9e-16;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:|||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518


RESULT 14
US-08-945-771-4
; Sequence 4, Application US/08945771
; Patent No. 6465623
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger J
;  APPLICANT: Sutherland, Robert L
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001700
;  CURRENT APPLICATION NUMBER: US/08/945,771
;  CURRENT FILING DATE: 1998-04-22
;  EARLIER APPLICATION NUMBER: PCT/US96/00258
;  EARLIER FILING DATE: 1996-MAY-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 621
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-08-945-771-4

  Query Match             44.0%;  Score 186;  DB 4;  Length 621;
  Best Local Similarity   54.1%;  Pred. No. 2.9e-16;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
```

```
                | |      :||||:|||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518


RESULT 15
US-08-472-595-9
; Sequence 9, Application US/08472595
; Patent No. 6001583
;  GENERAL INFORMATION:
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR TREATMENT
;    TITLE OF INVENTION:  OF BREAST CANCER
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,595
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-103
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  334 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-472-595-9

  Query Match           40.0%;  Score 169;  DB 3;  Length 334;
  Best Local Similarity  58.1%;  Pred. No. 2.2e-14;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db        272 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 331
```

```
Qy          73 SA 74
               :|
Db         332 AA 333


RESULT 16
US-08-207-575A-9
; Sequence 9, Application US/08207575A
; Patent No. 6037134
;  GENERAL INFORMATION:
;    APPLICANT:  Margolis, Benjamin L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR TREATMENT
;    TITLE OF INVENTION:  OF BREAST CANCER
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/207,575A
;      FILING DATE:  07-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  334 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-207-575A-9

  Query Match             40.0%;  Score 169;  DB 3;  Length 334;
  Best Local Similarity   58.1%;  Pred. No. 2.2e-14;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db         272 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 331

Qy          73 SA 74
```

```
                :|
Db        332 AA 333


RESULT 17
US-08-866-381A-1
; Sequence 1, Application US/08866381A
; Patent No. 6045797
;  GENERAL INFORMATION:
;    APPLICANT:  Ben Lewis Margolis
;    APPLICANT:  Joseph Schlessinger
;    TITLE OF INVENTION:  METHODS FOR TREATMENT OR DIAGNOSIS
;    TITLE OF INVENTION:  OF DISEASES OR CONDITIONS ASSOCIATED
;    TITLE OF INVENTION:  WITH A BLM DOMAIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,381A
;      FILING DATE:  May 30, 1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/212,234
;      FILING DATE:  March 14, 1994
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  226/043
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  335 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      OTHER INFORMATION:  BLM domain of GRB-7
US-08-866-381A-1
```

```
  Query Match             40.0%;  Score 169;  DB 3;  Length 335;
  Best Local Similarity   58.1%;  Pred. No. 2.2e-14;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db        273 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 332

Qy         73 SA 74
              :|
Db        333 AA 334


RESULT 18
US-09-280-598-51
; Sequence 51, Application US/09280598
; Patent No. 6391584
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnik, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  App, Harold
;    TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/280,598
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,820
;      FILING DATE:  02-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-067
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  51:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  335 amino acids
;      TYPE:  amino acid
```

```
;       TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-09-280-598-51

  Query Match             40.0%;  Score 169;  DB 4;  Length 335;
  Best Local Similarity   58.1%;  Pred. No. 2.2e-14;
  Matches   36;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||    : : || |
Db        273 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSLPTTCSGSSLS 332

Qy         73 SA 74
              :|
Db        333 AA 334


RESULT 19
US-08-866-381A-2
; Sequence 2, Application US/08866381A
; Patent No. 6045797
;  GENERAL INFORMATION:
;    APPLICANT:  Ben Lewis Margolis
;    APPLICANT:  Joseph Schlessinger
;    TITLE OF INVENTION:  METHODS FOR TREATMENT OR DIAGNOSIS
;    TITLE OF INVENTION:  OF DISEASES OR CONDITIONS ASSOCIATED
;    TITLE OF INVENTION:  WITH A BLM DOMAIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,381A
;      FILING DATE:  May 30, 1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/212,234
;      FILING DATE:  March 14, 1994
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  226/043
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
```

```
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  326 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FEATURE:
;       OTHER INFORMATION:  BLM domain of GRB-10
US-08-866-381A-2

  Query Match             39.5%;  Score 167;  DB 3;  Length 326;
  Best Local Similarity   57.1%;  Pred. No. 4e-14;
  Matches   40;  Conservative    5;  Mismatches   19;  Indels    6;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        252 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 309

Qy         63 HGSPTASSQS 72
                  ||||
Db        310 ----ILSSQS 315


RESULT 20
US-09-280-598-52
; Sequence 52, Application US/09280598
; Patent No. 6391584
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Skolnik, Edward Y.
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  App, Harold
;    TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;    TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
;    TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/280,598
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,820
```

```
;       FILING DATE:  02-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7683-067
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  52:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  326 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-09-280-598-52

  Query Match             39.5%;  Score 167;  DB 4;  Length 326;
  Best Local Similarity   57.1%;  Pred. No. 4e-14;
  Matches   40;  Conservative    5;  Mismatches   19;  Indels    6;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||  |||:|| || | |:||| |||  |  |:
Db        252 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 309

Qy         63 HGSPTASSQS 72
                    ||||
Db        310 ----ILSSQS 315


RESULT 21
US-09-023-905A-4
; Sequence 4, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
;  APPLICANT: Roberts, Thomas M.
;  APPLICANT:  King, Frederick J.
;  APPLICANT:  Harris, David F.
;  APPLICANT:  Hu, Erding
;  APPLICANT:  Spiegelman, Bruce
;  APPLICANT:  Chan, Joanne
;  TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
;  TITLE OF INVENTION:  Therefor
;  FILE REFERENCE: DFN-021
;  CURRENT APPLICATION NUMBER: US/09/023,905A
;  CURRENT FILING DATE:  1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/038,191
;  PRIOR FILING DATE: 1997-02-14
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1151
;   TYPE: PRT
;   ORGANISM: Danio rerio
US-09-023-905A-4

  Query Match             17.0%;  Score 72;  DB 4;  Length 1151;
```

```
  Best Local Similarity   28.6%;  Pred. No. 1.6;
  Matches   20;  Conservative   15;  Mismatches   33;  Indels    2;  Gaps    1;

Qy         14 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTAS--SQ 71
              :|:  : ||  :||  | |  :: ||: : |:     : |  ||:|   | |:    :|
Db        609 VRTSDQTSLHLVDFLVQNSGTLDRQTESGNAALHYCCTYEKPECLKLLLRGKPSIDLVNQ 668

Qy         72 SSATNMAIHR 81
              :  | : | |
Db        669 NGETALDIAR 678


RESULT 22
US-09-252-991A-28884
; Sequence 28884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 28884
;   LENGTH: 243
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28884

  Query Match             16.3%;  Score 69;  DB 4;  Length 243;
  Best Local Similarity   35.4%;  Pred. No. 0.4;
  Matches   23;  Conservative   10;  Mismatches   24;  Indels    8;  Gaps    3;

Qy         25 MDFSGQKSRVIENPTE-----ALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
              : |||  || ::          || :|:   |||:  | |||    || | : |  :: :|
Db         26 LPFSGASSRWLQRYAPALLAVALIIAMSISLAWQAAGWLRL--QRSPVAVAASPVSHESI 83

Qy         80 HRSQP 84
               || |
Db         84 -RSDP 87


RESULT 23
US-09-252-991A-19574
; Sequence 19574, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 19574
;   LENGTH: 863
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19574

  Query Match             15.6%;  Score 66;  DB 4;  Length 863;
  Best Local Similarity   31.2%;  Pred. No. 6.5;
  Matches   25;  Conservative    7;  Mismatches   38;  Indels   10;  Gaps    2;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
              || | ::  : | |:  |: ||| |  || |         ||  | ||   |:  |
Db        429 GRGGAAAVPVPPGRAAGEHGLVA-DRFGQPS---------LSARVIEGAGRRRLPCGTTD 478

Qy         62 THGSPTASSQSSATNMAIHR 81
                ||    |   |   : |
Db        479 RRESPYMQRQIFETEHNLFR 498


RESULT 24
US-09-198-452A-439
; Sequence 439, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides,
fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the
diagnosis, prevention
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 439
;   LENGTH: 653
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: 1...653
;   OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-439

  Query Match             15.4%;  Score 65;  DB 4;  Length 653;
  Best Local Similarity   31.3%;  Pred. No. 5.9;
  Matches   26;  Conservative   12;  Mismatches   35;  Indels   10;  Gaps    3;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
```

```
                 || |      ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db        358 GRKG------SPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG 411

Qy         62 THGSPTA---SSQSSATNMAIHR 81
                 |  |.     |::|    : ||
Db        412 -FGRKTGIELPSEASGLVPSPHR 433
```

RESULT 25
US-08-429-742-4
; Sequence 4, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,742
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-429-742-4

Query Match 14.5%; Score 61.5; DB 1; Length 388;
Best Local Similarity 32.8%; Pred. No. 8.1;
Matches 20; Conservative 11; Mismatches 27; Indels 3; Gaps 2;

```
Qy          3 RSGCSSQSI-SPMRSIS--ENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLR 59
              :|  |||::  |  ::|  ||| :      :|    ::|  : : |   ||| || | |
Db        228 QSSLSSQALQQPTSTVSMMENSSIPETDKEEKEHATQDPGLSTASAQHTGLARRKSGILL 287
```

```
Qy          60 L 60
               |
Db         288 L 288
```

RESULT 26
US-08-821-994-68
; Sequence 68, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Brassica napus
US-08-821-994-68

  Query Match   14.4%; Score 61; DB 3; Length 374;
  Best Local Similarity   31.8%; Pred. No. 9;
  Matches   21; Conservative   9; Mismatches   22; Indels   14; Gaps   2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL----GTHGSPTA 68
               |:| |::   | | :|   :          | | | : |||| :      || ||  |
Db         123 PVRRITKAKNVNMKYSAAVN----------DVEVPETVDWRKKGAVNAIKDQGTCGSCWA 172

Qy          69 SSQSSA 74
                | ::|
Db         173 FSTAAA 178
```

RESULT 27
US-09-252-991A-21729
; Sequence 21729, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

```
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 21729
;   LENGTH: 384
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21729

  Query Match             14.4%;  Score 61;  DB 4;  Length 384;
  Best Local Similarity   24.1%;  Pred. No. 9.3;
  Matches   21;  Conservative   13;  Mismatches   33;  Indels   20;  Gaps    3;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL----AWRKKGC 57
              |  || |||  |     :         |:: | :  |  :: | :|:|:      |:  |
Db         58 GCDGCRSQSSPPSGRADD---------GRRHRRVPRPPGSVPVGIEQGVRLMRMMRRLLC 108

Qy         58 LRLGTHGSPTASSQSSATNMAIHRSQP 84
                 |        : |:|  ||    :|
Db        109 WSAGL-------AMSAAVGMAAAADKP 128


RESULT 28
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
;  GENERAL INFORMATION:
;    APPLICANT:  Streuli, Michel
;    TITLE OF INVENTION:  No. 5994070el TRIO Molecules and Uses Related Thereto
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,267
;      FILING DATE:  1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/014,214
;      FILING DATE:  27 MARCH (1996)
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2860 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-826-267-2

  Query Match             14.3%;  Score 60.5;  DB 2;  Length 2860;
  Best Local Similarity   21.6%;  Pred. No. 2.1e+02;
  Matches   21;  Conservative   18;  Mismatches   45;  Indels   13;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQK--SRVIENPT-----------EALSVAVE 47
              :|  |  :  : |  :| ::||:  |    |  ||::  ||            |:   |:
Db       1800 EGEEGADAVPLPPPMAIQQHSLLQPDSQDDKASSRLLVRPTSSETPSAAELVSAIEELVK 1859

Qy         48 EGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
                :|   :    |   |  :: | : : |  :  | |
Db       1860 SKMALEDRPSSLLVDQGDSSSPSFNPSDNSLLSSSSP 1896


RESULT 29
US-09-252-991A-20992
; Sequence 20992, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 20992
;   LENGTH: 169
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20992

  Query Match             14.2%;  Score 60;  DB 4;  Length 169;
  Best Local Similarity   22.9%;  Pred. No. 3.8;
  Matches   27;  Conservative   11;  Mismatches   44;  Indels   36;  Gaps    4;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQ-KSRVIENPTEALSVA----VEEGLAWRKKG 56
              || || :      |  :    ||    |:  ||:     :| ::|      :| | |:
Db         19 GRLGCRASRSRARRHCANGQEVARSLPGRWPSRLGRCLFQAAAIAQGHRCGQGFAHRRAA 78

Qy         57 ------------CLRLGTHGSPTASSQS-------------------SATNMAIHRSQ 83
                         | ||| || | : :                        |   | ||\ |
Db         79 QTSNAAGSHRTQCGRLGVHGQPRSGASGHVQVERPGARRSRCALRARGARGPAAHRHQ 136


RESULT 30
```

US-09-252-991A-22999
; Sequence 22999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22999
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22999

Query Match 14.2%; Score 60; DB 4; Length 169;
Best Local Similarity 22.9%; Pred. No. 3.8;
Matches 27; Conservative 11; Mismatches 44; Indels 36; Gaps 4;

```
Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQ-KSRVIENPTEALSVA----VEEGLAWRKKG 56
              || || :      |  :    ||    |:  ||:     :| ::|      :| | |:
Db         19 GRLGCRASRSRARRHCANGQEVARSLPGRWPSRLGRCLFQAAAIAQGHRCGQGFAHRRAA 78

Qy         57 ------------CLRLGTHGSPTASSQS-------------------SATNMAIHRSQ 83
                        | ||| || | : :                        |   | || |
Db         79 QTSNAAGSHRTQCGRLGVHGQPRSGASGHVQVERPGARRSRCALRARGARGPAAHRHQ 136
```

RESULT 31
US-09-252-991A-25204
; Sequence 25204, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25204
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25204

```
  Query Match             14.2%;  Score 60;  DB 4;  Length 169;
  Best Local Similarity   22.9%;  Pred. No. 3.8;
  Matches   27;  Conservative   11;  Mismatches   44;  Indels   36;  Gaps    4;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQ-KSRVIENPTEALSVA----VEEGLAWRKKG 56
              || || :      |  :    ||    |:  ||:     :| ::|      :| | |:
Db         19 GRLGCRASRSRARRHCANGQEVARSLPGRWPSRLGRCLFQAAAIAQGHRCGQGFAHRRAA 78

Qy         57 ------------CLRLGTHGSPTASSQS------------------SATNMAIHRSQ 83
                        | ||| || | : :                      |   | || |
Db         79 QTSNAAGSHRTQCGRLGVHGQPRSGASGHVQVERPGARRSRCALRARGARGPAAHRHQ 136


RESULT 32
US-09-252-991A-26569
; Sequence 26569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 26569
;   LENGTH: 169
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26569

  Query Match             14.2%;  Score 60;  DB 4;  Length 169;
  Best Local Similarity   22.9%;  Pred. No. 3.8;
  Matches   27;  Conservative   11;  Mismatches   44;  Indels   36;  Gaps    4;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQ-KSRVIENPTEALSVA----VEEGLAWRKKG 56
              || || :      |  :    ||    |:  ||:     :| ::|      :| | |:
Db         19 GRLGCRASRSRARRHCANGQEVARSLPGRWPSRLGRCLFQAAAIAQGHRCGQGFAHRRAA 78

Qy         57 ------------CLRLGTHGSPTASSQS------------------SATNMAIHRSQ 83
                        | ||| || | : :                      |   | || |
Db         79 QTSNAAGSHRTQCGRLGVHGQPRSGASGHVQVERPGARRSRCALRARGARGPAAHRHQ 136


RESULT 33
US-09-252-991A-31908
; Sequence 31908, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
```

```
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 31908
;   LENGTH: 169
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31908

  Query Match             14.2%;  Score 60;  DB 4;  Length 169;
  Best Local Similarity   22.9%;  Pred. No. 3.8;
  Matches   27;  Conservative   11;  Mismatches   44;  Indels   36;  Gaps    4;

Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQ-KSRVIENPTEALSVA----VEEGLAWRKKG 56
              || || :      |  :    ||    |:  ||:     :| ::|      :| | |:
Db         19 GRLGCRASRSRARRHCANGQEVARSLPGRWPSRLGRCLFQAAAIAQGHRCGQGFAHRRAA 78

Qy         57 ------------CLRLGTHGSPTASSQS-------------------SATNMAIHRSQ 83
                        | ||| || | : :                        |   | || |
Db         79 QTSNAAGSHRTQCGRLGVHGQPRSGASGHVQVERPGARRSRCALRARGARGPAAHRHQ 136


RESULT 34
US-09-598-747-27
; Sequence 27, Application US/09598747
; Patent No. 6531648
; GENERAL INFORMATION:
;  APPLICANT:  Lanahan, Michael B.
;  APPLICANT:  Desai, Nalini M.
;  APPLICANT:  Gasdaska, Pamela Y.
;  TITLE OF INVENTION:  GRAIN PROCESSING METHOD AND TRANSGENIC PLANTS USEFUL
;  TITLE OF INVENTION:  THEREIN
;  FILE REFERENCE:  A-31383P1
;  CURRENT APPLICATION NUMBER: US/09/598,747
;  CURRENT FILING DATE:  2000-06-21
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-09-598-747-27

  Query Match             14.2%;  Score 60;  DB 4;  Length 310;
  Best Local Similarity   36.6%;  Pred. No. 9.3;
  Matches   15;  Conservative    8;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          6 CSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAV 46
              | :||:    ||   :: |:||| :  ||  : |  |: ||
```

```
Db         78 CRAQSLRFGTSIISETVTAVDFSARPFRVASDSTTVLADAV 118


RESULT 35
US-09-107-532A-6160
; Sequence 6160, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND
THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 6160:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 480 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...480
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6160:
US-09-107-532A-6160

  Query Match             14.2%;  Score 60;  DB 4;  Length 480;
  Best Local Similarity   35.6%;  Pred. No. 18;
```

```
  Matches   16;  Conservative    9;  Mismatches   18;  Indels    2;  Gaps    1;

Qy         22 LVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSP 66
              || :   |: :  : :|::|| || ||||    :   :  ||| |
Db        301 LVCLGVIGEIASWVTSPSKALHVAAEEGLL--PEYFAKENTHGVP 343


RESULT 36
US-09-328-352-4668
; Sequence 4668, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4668
;   LENGTH: 950
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-4668

  Query Match            14.2%;  Score 60;  DB 4;  Length 950;
  Best Local Similarity  28.6%;  Pred. No. 48;
  Matches   20;  Conservative   15;  Mismatches   33;  Indels    2;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK-GCLR 59
              :| :|  || : |   :: :: ||::  |::   :    |   : :|| |  ||  |  |
Db        390 KGTNG-KSQGVVPFLKVANDTAVAVNQGGKRKGAVCAYLETWHLDIEEFLELRKNTGDDR 448

Qy         60 LGTHGSPTAS 69
                ||   ||:
Db        449 RRTHDMNTAN 458


RESULT 37
US-09-107-532A-4552
; Sequence 4552, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND
THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
```

```
;              MEDIUM TYPE: CD/ROM ISO9660
;              COMPUTER: PC
;              OPERATING SYSTEM: <Unknown>
;              SOFTWARE: ASCII
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/107,532A
;              FILING DATE: 30-Jun-1998
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 60/085,598
;              FILING DATE: 14 May 1998
;              APPLICATION NUMBER: 60/051571
;              FILING DATE: July 2, 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ariniello, Pamela Deneke
;              REGISTRATION NUMBER: 40,489
;              REFERENCE/DOCKET NUMBER: GTC-012
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (781)893-5007
;              TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 4552:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1297 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         HYPOTHETICAL: YES
;         ORIGINAL SOURCE:
;              ORGANISM: Enterococcus faecium
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1...1297
;         SEQUENCE DESCRIPTION: SEQ ID NO: 4552:
US-09-107-532A-4552

  Query Match             14.1%;  Score 59.5;  DB 4;  Length 1297;
  Best Local Similarity   28.9%;  Pred. No. 89;
  Matches   22;  Conservative    9;  Mismatches   24;  Indels   21;  Gaps    4;

Qy          26 DFSGQKSRV-----------IENPTEALSV-----AVEEG----LAWRKKGCLRLGTHGS 65
               :|||  |                | : ::|      |:||      |   |     |:||
Db         402 EFSGNTSNAGFTHPVTYASDFNRPEDEVNVHYRYGEVKEGDNKATHWVGDGSSNNNTNGS 461

Qy          66 PTASSQSSATN-MAIH 80
               ||:  :||| | :| |
Db         462 PTSQEKSSAINTVAYH 477


RESULT 38
US-09-320-878-4
; Sequence 4, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
```

```
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-4

  Query Match             14.1%;  Score 59.5;  DB 3;  Length 1346;
  Best Local Similarity   34.6%;  Pred. No. 94;
  Matches   18;  Conservative    9;  Mismatches   14;  Indels   11;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
              |:| |  :|| |:||  : :|          | |:||  ||:  : : || |
Db        972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023


RESULT 39
US-09-141-908-5
; Sequence 5, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary
;  APPLICANT: MCDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
;  TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
;  FILE REFERENCE: 300622002100
;  CURRENT APPLICATION NUMBER: US/09/141,908
;  CURRENT FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: PROV. 60/076,919
;  EARLIER FILING DATE: 1998-03-05
;  EARLIER APPLICATION NUMBER: PROV. 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 31
```

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-141-908-5

  Query Match             14.1%;  Score 59.5;  DB 4;  Length 1346;
  Best Local Similarity   34.6%;  Pred. No. 94;
  Matches   18;  Conservative    9;  Mismatches   14;  Indels   11;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
              |:| |  :|| |:||  : :|          | |:||  ||:  : : || |
Db        972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023


RESULT 40
US-09-657-440-4
; Sequence 4, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/657,440
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: CIP OF 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-657-440-4

  Query Match             14.1%;  Score 59.5;  DB 4;  Length 1346;
  Best Local Similarity   34.6%;  Pred. No. 94;
  Matches   18;  Conservative    9;  Mismatches   14;  Indels   11;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
              |:| |  :|| |:||  : :|          | |:||  ||:  : : || |
Db        972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023


Search completed: January 13, 2004, 16:23:28
Job time : 20.8425 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:19:27 ; Search time 18.5197 Seconds
(without alignments)
436.194 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	191	45.2	535	2	C46243	epidermal growth f
2	189	44.7	548	2	I39175	SH2-domain protein
3	186	44.0	621	2	I49199	growth factor rece
4	179	42.3	532	2	JC5412	epidermal growth f
5	74.5	17.6	655	2	H96692	probable receptor
6	72.5	17.1	369	2	JQ2278	hydroxymethylbilan
7	70.5	16.7	346	2	AB3057	conserved hypothet
8	70.5	16.7	346	2	D98229	hypothetical prote
9	70	16.5	1520	2	T00273	hypothetical prote
10	69	16.3	235	2	S39652	secretion protein
11	68	16.1	653	2	A86543	transglycolase/tra
12	68	16.1	1240	2	T48800	SMT4 related prote
13	65	15.4	612	2	T32368	hypothetical prote

14	65	15.4	653	2	E72080	penicillin-binding
15	65	15.4	1791	2	T02345	hypothetical prote
16	64.5	15.2	196	2	C64891	ferripyochelin-bin
17	64.5	15.2	209	2	S44298	probable orotate p
18	64.5	15.2	256	2	T47860	transcription fact
19	64	15.1	313	2	F72575	hypothetical prote
20	64	15.1	470	1	S56565	hypothetical 53K p
21	64	15.1	470	2	F91291	probable regulator
22	64	15.1	470	2	H86132	probable regulator
23	63.5	15.0	2274	2	T30258	adenomatous polypo
24	63	14.9	197	2	T36696	probable regulator
25	62	14.7	242	2	AH1146	transcription regu
26	62	14.7	242	2	AH1505	transcription regu
27	62	14.7	559	2	T23571	hypothetical prote
28	62	14.7	665	2	T00015	unc-14 protein - C
29	62	14.7	1753	2	T00350	hypothetical prote
30	61.5	14.5	404	2	AB2188	hypothetical prote
31	61.5	14.5	785	2	T00474	hypothetical prote
32	61	14.4	324	2	E72536	probable oligopept
33	61	14.4	661	1	TNBE12	74K alpha trans-in
34	61	14.4	733	2	S44876	ZC21.4 protein - C
35	60.5	14.3	322	1	W2WLE	E2 protein - human
36	60.5	14.3	389	2	S53975	probable membrane
37	60.5	14.3	462	2	AE1323	3-isopropylmalate
38	60.5	14.3	512	2	AD0107	hypothetical prote
39	60.5	14.3	614	2	A32608	thyroid hormone re
40	60.5	14.3	716	2	T47449	hypothetical prote
41	60.5	14.3	803	2	A86655	hypothetical prote
42	60	14.2	263	2	JN0817	beta-hemolysin pre
43	60	14.2	460	2	B82549	hypothetical prote
44	60	14.2	548	2	T05671	hypothetical prote
45	60	14.2	692	2	T00025	PSD-95 binding pro

ALIGNMENTS

RESULT 1
C46243
epidermal growth factor-receptor-binding protein GRB-7 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: C46243
R;Margolis, B.; Silvennoinen, O.; Comoglio, F.; Roonprapunt, C.; Skolnik, E.; Ullrich, A.; Schlessinger, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 8894-8898, 1992
A;Title: High-efficiency expression/cloning of epidermal growth factor-receptor-binding proteins with Src homology 2 domains.
A;Reference number: A46243; MUID:93028373; PMID:1409582
A;Accession: C46243
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-535 <MAR>
A;Cross-references: GB:M94450; NID:g193619; PIDN:AAA37733.1; PID:g193620
A;Note: sequence extracted from NCBI backbone (NCBIP:115328)
C;Superfamily: pleckstrin repeat homology; SH2 homology
C;Keywords: growth factor receptor

F;434-530/Domain: SH2 homology <SH2B>

  Query Match 45.2%; Score 191; DB 2; Length 535;
  Best Local Similarity 59.7%; Pred. No. 7.3e-14;
  Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         423 S-LSAAIHRTQP 433
```

RESULT 2
I39175
SH2-domain protein Grb-IR - human
N;Alternate names: insulin receptor cytoplasmic tail-binding protein Grb-IR
C;Species: Homo sapiens (man)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999
C;Accession: I39175
R;Liu, F.; Roth, R.A.
Proc. Natl. Acad. Sci. U.S.A. 92, 10287-10291, 1995
A;Title: Grb-IR: a SH2-domain containing protein that binds to the insulin
receptor and inhibits its function.
A;Reference number: I39175; MUID:96036069; PMID:7479769
A;Accession: I39175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-548 <RES>
A;Cross-references: EMBL:U34355; NID:g1079573; PIDN:AAA88819.1; PID:g1079574
A;Note: cloned by a yeast two-hybrid screen with the insulin receptor
cytoplasmic domain as the bait
C;Genetics:
A;Gene: GDB:IRBP
A;Cross-references: GDB:697228
C;Superfamily: pleckstrin repeat homology; SH2 homology
F;447-541/Domain: SH2 homology <SH2B>

  Query Match 44.7%; Score 189; DB 2; Length 548;
  Best Local Similarity 53.0%; Pred. No. 1.3e-13;
  Matches 44; Conservative 11; Mismatches 26; Indels 2; Gaps 2;

```
Qy           1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
               | |    |   :|:||:||||||||||||||  |||||| || | |:||| ||||:   |:
Db         365 QQRKALLSPFSTPVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKRS-TRM 423

Qy          61 GTHGSPTASSQSSATNMAIHRSQ 83
                  || :    |: : : |||:|
Db         424 NILGSQSPLHPSTLSTV-IHRTQ 445
```

RESULT 3
I49199
growth factor receptor binding protein Grb10 - mouse
C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I49199
R;Ooi, J.; Yajnik, V.; Immanuel, D.; Gordon, M.; Moskow, J.J.; Buchberg, A.M.; Margolis, B.
Oncogene 10, 1621-1630, 1995
A;Title: The cloning of Grb10 reveals a new family of SH2 domain proteins.
A;Reference number: I49199; MUID:95249278; PMID:7731717
A;Accession: I49199
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-621 <RES>
A;Cross-references: EMBL:U18996; NID:g841209; PIDN:AAB53687.1; PID:g841210
C;Genetics:
A;Gene: Grb10
C;Superfamily: pleckstrin repeat homology; SH2 homology
C;Keywords: growth factor receptor
F;520-614/Domain: SH2 homology <SH2B>

  Query Match 44.0%; Score 186; DB 2; Length 621;
  Best Local Similarity 54.1%; Pred. No. 3.2e-13;
  Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                      ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518
```

RESULT 4
JC5412
epidermal growth factor receptor-binding protein GRB-7 - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 21-Jul-2000
C;Accession: JC5412
R;Kishi, T.; Sasaki, H.; Akiyama, N.; Ishizuka, T.; Sakamoto, H.; Aizawa, S.; Sugimura, T.; Terada, M.
Biochem. Biophys. Res. Commun. 232, 5-9, 1997
A;Title: Molecular cloning of human GRB-7 co-amplified with CAB1 and c-ERBB-2 in primary gastric cancer.
A;Reference number: JC5412; MUID:97236270; PMID:9125150
A;Accession: JC5412
A;Molecule type: mRNA
A;Residues: 1-532 <KIS>
A;Cross-references: DDBJ:D43772; NID:g601890; PIDN:BAA07827.1; PID:g601891
C;Comment: This protein contains a pleckstrin domain which mediates protein-protein interaction during signal transduction.
C;Genetics:
A;Gene: GDB:GRB7
A;Cross-references: GDB:1297554; OMIM:601522
C;Superfamily: pleckstrin repeat homology
F;231-336/Domain: pleckstrin #status predicted <PLE>
F;432-532/Domain: SH2 #status predicted <SH2>

  Query Match 42.3%; Score 179; DB 2; Length 532;

```
  Best Local Similarity   59.2%;  Pred. No. 1.7e-12;
  Matches   42;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| ||||||:||  |||||   ||     | :| :
Db        363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 419

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db        420 S-LSAAIHRTQ 429


RESULT 5
H96692
probable receptor serine/threonine kinase PR5K T4O24.8 [imported] - Arabidopsis
thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96692
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;
Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96692
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AE005173; NID:g11128390; PIDN:AAG31195.1; GSPDB:GN00141
C;Genetics:
A;Gene: T4O24.8
A;Map position: 1

  Query Match             17.6%;  Score 74.5;  DB 2;  Length 655;
  Best Local Similarity   25.6%;  Pred. No. 2.2;
  Matches   23;  Conservative   16;  Mismatches   34;  Indels   17;  Gaps    3;

Qy         11 ISPMRSISENSLVAMDFSGQKSRVIENP----------TEALSVAVEEGLAWRKKG 56
              : |   :  || :  || |  || : ||             |: | :|:|:| |
Db        164 LPPSLKLEGNSFLLNDFGGSCSRNVSNPASRTALNTLESTPSTDNLKIALEDGFALEVNS 223

Qy         57 CLR--LGTHGSPTASSQSSATNMAIHRSQP 84
                |  : : |:    ||:|:  :  :| :|
```

Db 224 DCRTCIDSKGA-CGFSQTSSRFVCYYRQEP 252

RESULT 6
JQ2278
hydroxymethylbilane synthase (EC 4.3.1.8) precursor, chloroplast - garden pea
N;Alternate names: porphobilinogen deaminase
C;Species: Pisum sativum (garden pea)
C;Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 16-Jul-1999
C;Accession: S35873; JQ2278; PQ0748; S13475
R;Smith, A.G.
submitted to the EMBL Data Library, June 1993
A;Reference number: S35873
A;Accession: S35873
A;Molecule type: mRNA
A;Residues: 1-369 <SMI>
A;Cross-references: EMBL:X73418; NID:g313723; PIDN:CAA51820.1; PID:g313724
R;Witty, M.; Wallace-Cook, A.D.M.; Albrecht, H.; Spano, A.J.; Michel, H.;
Shabanowitz, J.; Hunt, D.F.; Timko, M.P.; Smith, A.G.
Plant Physiol. 103, 139-147, 1993
A;Title: Structure and expression of chloroplast-localized porphobilinogen
deaminase from pea (Pisum sativum L.) isolated by redundant polymerase chain
reaction.
A;Reference number: JQ2278; MUID:94269188; PMID:7516080
A;Accession: JQ2278
A;Molecule type: DNA
A;Residues: 1-369 <WIT>
A;Cross-references: GB:X73418; NID:g313723; PIDN:CAA51820.1; PID:g313724
A;Accession: PQ0748
A;Molecule type: protein
A;Residues: 47-63;64,109-119;125-143;144,167-172;219-226;227,275-286;323-
332;339-349 <WI2>
R;Spano, A.J.; Timko, M.P.
Biochim. Biophys. Acta 1076, 29-36, 1991
A;Title: Isolation, characterization and partial amino acid sequence of a
chloroplast-localized porphobilinogen deaminase from pea (Pisum sativum L.).
A;Reference number: S13475; MUID:91098265; PMID:1986793
A;Accession: S13475
A;Molecule type: protein
A;Residues: 47-56,'DX',59-60,'G' <SPA>
A;Note: 9-Cys and 11-Gln were also found
C;Comment: This enzyme catalyzes the polymerization of four porphobilinogen
monopyrrole units into the linear tetrapyrrole hydroxymethylbilane necessary for
the formation of chlorophyll and heme in plant cells.
C;Genetics:
A;Genome: nuclear
A;Introns: 204/3; 273/3; 333/1
C;Superfamily: hydroxymethylbilane synthase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase; chlorophyll biosynthesis;
chloroplast; porphyrin biosynthesis
F;1-46/Domain: transit peptide (chloroplast) #status predicted <SIG>
F;47-369/Product: hydroxymethylbilane synthase #status experimental <MAT>
F;303/Modified site: dipyrrolylmethanemethyl (Cys) (covalent) #status predicted

Query Match 17.1%; Score 72.5; DB 2; Length 369;
Best Local Similarity 33.3%; Pred. No. 1.9;
Matches 27; Conservative 9; Mismatches 38; Indels 7; Gaps 2;

```
Qy          7 SSQSISPMRSISENSL----VAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              || | |   : |  ||     :  ||  |:          |:|||:     |   :|:||
Db          7 SSSSFSLPSAPSNPSLSLFTSSFRFSSFKTSPFSKCRIRASLAVEQQTQQNKTALIRIGT 66

Qy         63 HGSPTASSQSSATN---MAIH 80
                |||  | :|:  |    || |
Db         67 RGSPLALAQAHETRDKLMASH 87
```

RESULT 7
AB3057
conserved hypothetical protein Atu4071 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 06-Jan-2003
C;Accession: AB3057
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClelland, E.; Palmieri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB3057
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44872.1; PID:g17742520; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4071
A;Map position: linear chromosome
C;Superfamily: uncharacterized conserved protein

  Query Match   16.7%; Score 70.5; DB 2; Length 346;
  Best Local Similarity   27.9%; Pred. No. 2.9;
  Matches   24; Conservative   13; Mismatches   30; Indels   19; Gaps   4;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA--------WRK 54
              |:||    : |  |     |::  :: |  | :|    || :||| ||         | |
Db        194 RAGCDLNPLDPSSSEDRLRLMSYIWADQTDR-LERTAAALRIAVENGLQVEKADAVDWLK 252

Qy         55 KGCLRLGTHGSPTASSQSSATNMAIH 80
               :   ||       |:  : ||::  |
Db        253 R---RL-------ATQHTGATHVVYH 268
```

RESULT 8
D98229

hypothetical protein AGR_L_1570 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 06-Jan-2003
C;Accession: D98229
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmiel, K.; Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gurson, J.; Lomo, C.; Sear, C.; Strub, G.; Cielo, C.; Slater, S.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58.
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D98229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89358.1; PID:g15159206; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1570
A;Map position: linear chromosome
C;Superfamily: uncharacterized conserved protein

  Query Match 16.7%; Score 70.5; DB 2; Length 346;
  Best Local Similarity 27.9%; Pred. No. 2.9;
  Matches 24; Conservative 13; Mismatches 30; Indels 19; Gaps 4;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA--------WRK 54
              |:||    : |  |     |::  :: |  | :|    || :||| ||         | |
Db        194 RAGCDLNPLDPSSSEDRLRLMSYIWADQTDR-LERTAAALRIAVENGLQVEKADAVDWLK 252

Qy         55 KGCLRLGTHGSPTASSQSSATNMAIH 80
              :   ||       |:  : ||::  |
Db        253 R---RL-------ATQHTGATHVVYH 268
```

RESULT 9
T00273
hypothetical protein KIAA0595 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00273
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.
A;Reference number: Z14086; MUID:98290545; PMID:9628581
A;Accession: T00273
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1520 <NAG>
A;Cross-references: EMBL:AB011167; NID:g3043713; PIDN:BAA25521.1; PID:g3043714
A;Experimental source: brain
C;Genetics:

A;Note: KIAA0595

  Query Match 16.5%; Score 70; DB 2; Length 1520;
  Best Local Similarity 25.8%; Pred. No. 19;
  Matches 24; Conservative 17; Mismatches 32; Indels 20; Gaps 3;

```
Qy          1 QGRSGCSSQSISP----MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG 56
              ||| | :|:|:|         | :|  :   |  :|| :  | :           ||:
Db       1276 QGRRGRNSRSVSSGSNRTSEASSSSSSSSSSRSRSRSLSPPHK----------RWRRSS 1325

Qy         57 C------LRLGTHGSPTASSQSSATNMAIHRSQ 83
              |       |  :  | ::|| ||::: :  ||:
Db       1326 CSSSGRSRRCSSSSSSSSSSSSSSSSSSSRSR 1358
```

RESULT 10
S39652
secretion protein XcpP PA3104 [imported] - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2000
C;Accession: S39652; H83258
R;Akrim, M.; Bally, M.; Ball, G.; Tommassen, J.; Teerink, H.; Filloux, A.; Lazdunski, A.
Mol. Microbiol. 10, 431-443, 1993
A;Title: Xcp-mediated protein secretion in Pseudomonas aeruginosa: identification of two additional genes and evidence for regulation of xcp gene expression.
A;Reference number: S39652; MUID:95020542; PMID:7934833
A;Accession: S39652
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-235 <AKR>
A;Cross-references: EMBL:X68594; NID:g431183; PIDN:CAA48581.1; PID:g431184
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83258
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-235 <STO>
A;Cross-references: GB:AE004734; GB:AE004091; NID:g9949204; PIDN:AAG06492.1; GSPDB:GN00131; PASP:PA3104
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: xcpP; PA3104

  Query Match 16.3%; Score 69; DB 2; Length 235;
  Best Local Similarity 35.4%; Pred. No. 2.7;
  Matches 23; Conservative 10; Mismatches 24; Indels 8; Gaps 3;

```
Qy          25 MDFSGQKSRVIENPTE-----ALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
               : |||  || ::          || :|:   |||:  | |||    || | : |  :: :|
Db          18 LPFSGASSRWLQRYAPALLAVALIIAMSISLAWQAAGWLRL--QRSPVAVAASPVSHESI 75

Qy          80 HRSQP 84
                || |
Db          76 -RSDP 79
```

RESULT 11
A86543
transglycolase/transpeptidase [imported] - Chlamydophila pneumoniae (strain J138)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 24-Aug-2001
C;Accession: A86543
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishii, K.; Hattori, M.; Kuhara, S.; Nakazawa, T.
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: A86543
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-653 <STO>
A;Cross-references: GB:BA000008; NID:g8978791; PIDN:BAA98627.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: pbp3
C;Superfamily: penicillin-binding protein 3

Query Match 16.1%; Score 68; DB 2; Length 653;
Best Local Similarity 30.5%; Pred. No. 12;
Matches 25; Conservative 14; Mismatches 39; Indels 4; Gaps 2;

```
Qy           3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
               |: |  :  ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db         353 RTLCPGRKGSPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG- 411

Qy          63 HGSPTA---SSQSSATNMAIHR 81
                |  |     |::|    : ||
Db         412 FGRKTGIELPSEASGLVPSPHR 433
```

RESULT 12
T48800
SMT4 related protein [imported] - Neurospora crassa
N;Alternate names: protein 15E6.80
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C;Accession: T48800
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, G.; Mewes, H.W.; Mannhaupt, G.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24541
A;Accession: T48800
A;Status: preliminary

```
A;Molecule type: DNA
A;Residues: 1-1240 <SCH>
A;Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.80
A;Experimental source: cosmid contig 15E6; strain 74
C;Genetics:
A;Gene: NCSP:15E6.80
A;Map position: 2
A;Introns: 8/3; 358/2

  Query Match             16.1%;  Score 68;  DB 2;  Length 1240;
  Best Local Similarity   34.4%;  Pred. No. 26;
  Matches   22;  Conservative    5;  Mismatches   23;  Indels   14;  Gaps    2;

Qy         32 SRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTA----------SSQSSATNMA 78
              |||    | || |     :|:   | :   | |||||             ||| : :|
Db        386 SRVTRT-TSALDVEGSRNMAFEPAGLIAQATAGSPTASTRRRPRLVDTLLSSQQALSNQY 444

Qy         79 IHRS 82
              |||
Db        445 EHRS 448


RESULT 13
T32368
hypothetical protein C01B12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C;Accession: T32368
R;Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid C01B12.
A;Reference number: Z21156
A;Accession: T32368
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-612 <SCH>
A;Cross-references: EMBL:AF025458; PIDN:AAB70976.1; GSPDB:GN00020; CESP:C01B12.3
A;Experimental source: strain Bristol N2; clone C01B12
C;Genetics:
A;Gene: CESP:C01B12.3
A;Map position: 2
A;Introns: 25/3; 60/2; 105/2; 138/3; 212/3; 319/3; 369/2; 467/2; 508/3; 573/1
C;Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

  Query Match             15.4%;  Score 65;  DB 2;  Length 612;
  Best Local Similarity   28.7%;  Pred. No. 25;
  Matches   29;  Conservative    8;  Mismatches   30;  Indels   34;  Gaps    4;

Qy         10 SISPMRSISE-------NSLVAMDFSGQKSRVIENPT--------------------EAL 42
              |  |   : |       || |  :  | | | ::|||                     |:|
Db        496 SSMPQTQLEEMLKNKNFNSPVKYNTDGMKDRELQNPTPITDHIDLPLHVASSQSWFNESL 555

Qy         43 SVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQ 83
              |  ||  | ||          || :|  ||   |:| ||:
Db        556 PVIKEEEEAKRKSNT----DTESPKSSKHSS---MSIRRSE 589
```

RESULT 14
E72080
penicillin-binding protein CP0335 [imported] - Chlamydophila pneumoniae (strains CWL029 and AR39)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C;Accession: E72080; A81588
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: E72080
A;Molecule type: DNA
A;Residues: 1-653 <ARN>
A;Cross-references: GB:AE001625; GB:AE001363; NID:g4376695; PIDN:AAD18563.1; PID:g4376700
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, E.K.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.; Weidman, J.; Khouri, H.; Craven, B.; Bowman, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; Eisen, J.; Fraser, C.M.
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: A81588
A;Molecule type: DNA
A;Residues: 1-653 <REA>
A;Cross-references: GB:AE002196; GB:AE002161; NID:g7189258; PIDN:AAF38189.1; PID:g7189263; GSPDB:GN00122; TIGR:CP0335
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: pbp3; CP0335
C;Superfamily: penicillin-binding protein 3

  Query Match 15.4%; Score 65; DB 2; Length 653;
  Best Local Similarity 31.3%; Pred. No. 27;
  Matches 26; Conservative 12; Mismatches 35; Indels 10; Gaps 3;

```
Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
              || |      ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db        358 GRKG------SPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG 411

Qy         62 THGSPTA---SSQSSATNMAIHR 81
                |  |     |::|    : ||
Db        412 -FGRKTGIELPSEASGLVPSPHR 433
```

RESULT 15
T02345
hypothetical protein KIAA0324 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02345

R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.; Jones, M.; Buckingham, J.; Chasteen, L.; Thompson, S.; Goodwin, L.; Bryant, J.; Tesmer, J.; Meincke, L.; Longmire, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A;Description: Sequencing of human chromosome 16p13.3.
A;Reference number: Z14664
A;Accession: T02345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1791 <RIC>
A;Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C;Genetics:
A;Map position: 16
A;Introns: 1610/2; 1706/2
A;Note: KIAA0324

Query Match 15.4%; Score 65; DB 2; Length 1791;
Best Local Similarity 28.1%; Pred. No. 89;
Matches 27; Conservative 12; Mismatches 35; Indels 22; Gaps 2;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKS----------------RVIENPTEALSVAV 46
              ||  || | |   | | :|  :   ||  |                 : : :|| |   ||
Db       1563 RSSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 1622

Qy         47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
               ||       | ||:      :  | ::|| ||:::
Db       1623 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 1658
```

RESULT 16
C64891
ferripyochelin-binding protein homolog b1400 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: C64891
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64891
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-196 <BLAT>
A;Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74482.1; PID:g1787667; UWGP:b1400
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: ferripyochelin binding protein

Query Match 15.2%; Score 64.5; DB 2; Length 196;
Best Local Similarity 28.0%; Pred. No. 7.2;
Matches 21; Conservative 15; Mismatches 24; Indels 15; Gaps 4;

```
Qy         17 ISENSLV-AMDFSGQKSR------VIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTAS 69
              | |||:| |  |   |:       :: :| :|:    |: |||:|:     |||
```

```
Db        109 IGENSIVGASAFVKAKAEMPANYLIVGSPAKAIRELSEQELAWKKQ-----GTHEYQVLV 163

Qy         70 SQSSATNMAIHRSQP 84
              ::    |    :|: :|
Db        164 TRCKQT---LHQVEP 175
```

RESULT 17
S44298
probable orotate phosphoribosyltransferase (EC 2.4.2.10) [similarity] - Coxiella burnetii
N;Alternate names: protein 209
C;Species: Coxiella burnetii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Mar-2000
C;Accession: S44298
R;Thiele, D.; Willems, H.; Oswald, W.; Krauss, H.
submitted to the EMBL Data Library, May 1994
A;Reference number: S44297
A;Accession: S44298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <THI>
A;Cross-references: EMBL:X79075; NID:g483518; PIDN:CAA55676.1; PID:g483520
C;Superfamily: orotate phosphoribosyltransferase; orotate phosphoribosyltransferase homology
C;Keywords: glycosyltransferase; pentosyltransferase
F;1-196/Domain: orotate phosphoribosyltransferase homology <OPT>

  Query Match   15.2%; Score 64.5; DB 2; Length 209;
  Best Local Similarity   25.4%; Pred. No. 7.8;
  Matches   17; Conservative   14; Mismatches   31; Indels   5; Gaps   1;

```
Qy         19 ENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG-----CLRLGTHGSPTASSQSS 73
              :| :      ||:: ::|:       |:  ||| |:||     |: : ::  | |    |
Db        105 QNQIEGRIRKGQRALIVEDLISTGKSALAAGLALREKGVTVTDCIAIFSYQLPQAQQNFS 164

Qy         74 ATNMAIH 80
                |:  |
Db        165 DANINCH 171
```

RESULT 18
T47860
transcription factor-like protein - Arabidopsis thaliana
N;Alternate names: protein T8B10.150
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47860
R;Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat M.Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24478
A;Accession: T47860
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-256 <RIE>
A;Cross-references: EMBL:AL138646

A;Experimental source: cultivar Columbia; BAC clone T8B10
C;Genetics:
A;Map position: 3
A;Note: T8B10.150

  Query Match 15.2%; Score 64.5; DB 2; Length 256;
  Best Local Similarity 32.6%; Pred. No. 10;
  Matches 29; Conservative 10; Mismatches 31; Indels 19; Gaps 4;

```
Qy          7 SSQSI----SPMRSISENSLVAMDFSGQKSRVI------ENPTEALSVAVEEGLAW---- 52
              || |:    |   | |: |||  : || | |        :|||    | :     |
Db         27 SSSSVVTSSSDSWSTSKRSLVQDNDSGGKRRKSNVSDDNKNPTSYRGVRMRSWGKWVSEI 86

Qy         53 ---RKKGCLRLGTHGSPTASSQSSATNMA 78
                 |||  : |||:  |||   : | ::|
Db         87 REPRKKSRIWLGTY--PTAEMAARAHDVA 113
```

RESULT 19
F72575
hypothetical protein APE1886 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: F72575
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: F72575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80891.1; PID:g5105578
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1886
C;Superfamily: Aeropyrum pernix hypothetical protein APE1886

  Query Match 15.1%; Score 64; DB 2; Length 313;
  Best Local Similarity 30.4%; Pred. No. 14;
  Matches 17; Conservative 10; Mismatches 21; Indels 8; Gaps 2;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEA-LSVAVEEGLAWRKK 55
              :|  |||: |    |       :  : |  : |   | :: |  |||:|||| ::
Db          3 RGPGGCSTTSYQSWRE-------SRSWRGAAAAVHSTPQQSRLEEAVEKGLAWARR 51
```

RESULT 20
S56565
hypothetical 53K protein (iadA-mcrD intergenic region) - Escherichia coli (strain K-12)
N;Alternate names: hypothetical protein f470

C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: S56565; F65248
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region
from 92.8 through 100 minutes.
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56565
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-470 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97236.1; PID:g537181
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: F65248
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-470 <BLAT>
A;Cross-references: GB:AE000504; GB:U00096; NID:g1790789; PIDN:AAC77296.1;
PID:g1790797; UWGP:b4340
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yjiR
C;Superfamily: hypothetical protein b1439

  Query Match 15.1%; Score 64; DB 1; Length 470;
  Best Local Similarity 26.3%; Pred. No. 23;
  Matches 25; Conservative 14; Mismatches 34; Indels 22; Gaps 4;

```
Qy          4 SGC-SSQSISPMRSISENSLVAMD----------FSGQKSRVIENPT--------EALSV 44
              ||| :| |:: |       :||::            |   :||| ||        ||| :
Db        175 SGCHNSMSLALMAVCKPGDIVAVESPCYYGSMQMLRGMGVKVIEIPTDPETGISVEALEL 234

Qy         45 AVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
              |:|:   |  || : :    :|       |   |:
Db        235 ALEQ---WPIKGIILVPNCNNPLGFIMPDARKRAV 266
```

RESULT 21
F91291
probable regulator [imported] - Escherichia coli (strain O157:H7, substrain RIMD
0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
C;Accession: F91291
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.;
Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida,
T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara,
S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7
and genomic comparison with a laboratory strain K-12.
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F91291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-470 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38725.1; PID:g13364780; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5302
C;Superfamily: hypothetical protein b1439

  Query Match   15.1%; Score 64; DB 2; Length 470;
  Best Local Similarity 26.3%; Pred. No. 23;
  Matches 25; Conservative 14; Mismatches 34; Indels 22; Gaps 4;

Qy   4 SGC-SSQSISPMRSISENSLVAMD----------FSGQKSRVIENPT--------EALSV 44
              ||| :| |:: | :||:: | :||| || ||| :
Db   175 SGCHNSMSLALMAVCKPGDIVAVESPCYYGSMQMLRGMGVKVIEIPTDPETGISVEALEL 234

Qy   45 AVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
              |:|: | || : :   :| | |:
Db   235 ALEQ---WPIKGIILVPNCNNPLGFIMPDARKRAV 266

RESULT 22
H86132
probable regulator yjiR [imported] - Escherichia coli (strain O157:H7, substrain
EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: H86132
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose,
D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.;
Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis,
N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.;
Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H86132
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-470 <STO>
A;Cross-references: GB:AE005174; NID:g12519358; PIDN:AAG59524.1; GSPDB:GN00145;
UWGP:Z5941
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yjiR

  Query Match   15.1%; Score 64; DB 2; Length 470;
  Best Local Similarity 26.3%; Pred. No. 23;
  Matches 25; Conservative 14; Mismatches 34; Indels 22; Gaps 4;

Qy   4 SGC-SSQSISPMRSISENSLVAMD----------FSGQKSRVIENPT--------EALSV 44
              ||| :| |:: | :||:: | :||| || ||| :

```
Db        175 SGCHNSMSLALMAVCKPGDIVAVESPCYYGSMQMLRGMGVKVIEIPTDPETGISVEALEL 234

Qy         45 AVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
              |:|:   |  || : :    :|       |   |:
Db        235 ALEQ---WPIKGIILVPNCNNPLGFIMPDARKRAV 266
```

RESULT 23
T30258
adenomatous polyposis coli protein 2 - mouse
N;Alternate names: APC2 protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30258
R;van Es, J.H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Miles, A.; Kuipers, J.; Destree, O.; Peifer, M.; Clevers, H.
Curr. Biol. 9, 105-108, 1999
A;Title: Identification of APC2, a homoloque of the adenomatous polyposis coli tumour suppressor.
A;Reference number: Z20796; MUID:99147086; PMID:10021369
A;Accession: T30258
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2274 <VAN>
A;Cross-references: EMBL:AJ130783; NID:g4210431; PIDN:CAA10207.1; PID:g4210432
C;Genetics:
A;Gene: APC2
A;Introns: 47/3; 78/1; 138/2; 174/3; 212/3; 238/3; 271/3; 396/1; 428/1; 474/3; 500/3; 539/3; 611/2

  Query Match   15.0%; Score 63.5; DB 2; Length 2274;
  Best Local Similarity   30.2%; Pred. No. 1.7e+02;
  Matches   19; Conservative   6; Mismatches   19; Indels   19; Gaps   2;

```
Qy         41 ALSVAVEEGLAWRKKGCL----------------RLGTHGSPTASSQSSATNMAIHR--- 81
              |:| : |  :| |: |||                | |  |:| |       | |:|
Db        296 AMSSSPESCVAMRRSGCLPLLLQILHGTEAGSVGRAGIPGAPGAKDARMRANAALHNIVF 355

Qy         82 SQP 84
              |||
Db        356 SQP 358
```

RESULT 24
T36696
probable regulatory protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36696
R;Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21597
A;Accession: T36696
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-197 <MUR>

A;Cross-references: EMBL:AL049731; PIDN:CAB41735.1; GSPDB:GN00070;
SCOEDB:SCH66.08c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCH66.08c

Query Match 14.9%; Score 63; DB 2; Length 197;
Best Local Similarity 40.5%; Pred. No. 11;
Matches 15; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

```
Qy         38 PTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSA 74
              |    | ::| | ||||||  |:       ||  | |
Db          3 PRGLASCSLEPGAAWRKKGWARITVRDIAAASGVSMA 39
```

RESULT 25
AH1146
transcription regulator GntR family homolog lmo0575 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
C;Accession: AH1146
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1146
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98654.1; PID:g16409951; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0575
C;Superfamily: transcription regulator GntR

Query Match 14.7%; Score 62; DB 2; Length 242;
Best Local Similarity 36.4%; Pred. No. 18;
Matches 16; Conservative 6; Mismatches 16; Indels 6; Gaps 1;

```
Qy         40 EALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQ 83
              :|| | | ||| :||:|      ||:    |   | : ||  :
Db         49 KALEVLVLEGLLYRKRG------HGTFIIKSALDADRLQIHNQE 86
```

RESULT 26
AH1505

transcription regulator GntR family homolog lin0584 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
C;Accession: AH1505
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95816.1; PID:g16413024; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0584
C;Superfamily: transcription regulator GntR

Query Match 14.7%; Score 62; DB 2; Length 242;
Best Local Similarity 36.4%; Pred. No. 18;
Matches 16; Conservative 6; Mismatches 16; Indels 6; Gaps 1;

```
Qy         40 EALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQ 83
              :|| | | ||| :||:|      ||:    |   | : ||  :
Db         49 KALEVLVLEGLLYRKRG------HGTFIIKSALDADRLQIHNQE 86
```

RESULT 27
T23571
hypothetical protein K10D3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23571
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19762
A;Accession: T23571
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-559 <WIL>
A;Cross-references: EMBL:Z75545; PIDN:CAA99884.1; GSPDB:GN00019; CESP:K10D3.2
A;Experimental source: clone K10D3
C;Genetics:
A;Gene: CESP:K10D3.2
A;Map position: 1
A;Introns: 210/3; 249/3; 277/2; 337/2; 371/2; 419/2; 479/2

```
  Query Match             14.7%;  Score 62;  DB 2;  Length 559;
  Best Local Similarity   23.8%;  Pred. No. 49;
  Matches   25;  Conservative   16;  Mismatches   34;  Indels   30;  Gaps    4;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL---AWRKKGCLRLGTH 63
              |  : || |    |||:|   ||  :| ::     :  | || :    ||:    |  :
Db        111 SDSARSPNR---PNSLIANFVSGDATRFVDVNDNEIREANEEIIRKDRWRRDSARRCSSG 167

Qy         64 G----------------SPTASSQSSATN--------MAIHRSQP 84
              |                :||: : :|: |        :|:|   |
Db        168 GQNQKRTFADILEKNVTAPTSMAITSSDNEKPPKLDFLAMHHEMP 212


RESULT 28
T00015
unc-14 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C;Accession: T00015
R;Ogura, K.; Shirakawa, M.; Thomas, B.M.; Siegfried, H.; Yasumi, O.
Genes Dev. 11, 1801-1811, 1997
A;Title: The UNC-14 protein required for axonal elongation and guidance in
Caenorhabditis elegans interacts with the serine / threonine kinase UNC-51.
A;Reference number: Z14053; MUID:97384993; PMID:9242488
A;Accession: T00015
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-665 <OGU>
A;Cross-references: EMBL:AB000913; NID:g2308978; PIDN:BAA21715.1; PID:g2308979
A;Experimental source: strain N2
C;Genetics:
A;Gene: unc-14
A;Map position: I

  Query Match             14.7%;  Score 62;  DB 2;  Length 665;
  Best Local Similarity   23.8%;  Pred. No. 60;
  Matches   25;  Conservative   16;  Mismatches   34;  Indels   30;  Gaps    4;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL---AWRKKGCLRLGTH 63
              |  : || |    |||:|   ||  :| ::     :  | || :    ||:    |  :
Db        111 SDSARSPNR---PNSLIANFVSGDATRFVDVNDNEIREANEEIIRKDRWRRDSARRCSSG 167

Qy         64 G----------------SPTASSQSSATN--------MAIHRSQP 84
              |                :||: : :|: |        :|:|   |
Db        168 GQNQKRTFADILEKNVTAPTSMAITSSDNEKPPKLDFLAMHHEMP 212


RESULT 29
T00350
hypothetical protein KIAA0708 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00350
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.;
Nomura, N.; Ohara, O.
DNA Res. 5, 169-176, 1998
```

A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.
A;Reference number: Z14142; MUID:98403880; PMID:9734811
A;Accession: T00350
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1753 <ISH>
A;Cross-references: EMBL:AB014608; NID:g3327229; PIDN:BAA31683.1; PID:g3327230
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0708

  Query Match 14.7%; Score 62; DB 2; Length 1753;
  Best Local Similarity 32.9%; Pred. No. 1.9e+02;
  Matches 24; Conservative 9; Mismatches 28; Indels 12; Gaps 4;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRKKGCLRLG 61
              |: | |  :  :| | |  |  :  | |  || :::|      :||   ||  ||    |
Db       1637 RADCLSTGMELLRRIQERLLAILQHSAQDFRVGLQSP------SVE---AWEAKGPNMPG 1687

Qy         62 THGSPTASSQSSA 74
              :   | |||   |
Db       1688 S--QPQASSGPEA 1698
```

RESULT 30
AB2188
hypothetical protein alr3057 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2188
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.; Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120.
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2188
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74756.1; PID:g17132151; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3057

  Query Match 14.5%; Score 61.5; DB 2; Length 404;
  Best Local Similarity 33.3%; Pred. No. 38;
  Matches 17; Conservative 10; Mismatches 13; Indels 11; Gaps 2;

```
Qy         15 RSIS---------ENSLVAMDFSGQKSRVIENP--TEALSVAVEEGLAWRK 54
              ||:|         | || ||:: |:|:  | |   |:  :||  : : ||:
Db         95 RSLSSDFMHFHRLEPSLAAMNWQGEKTIFIHNDIHTQMATVADRKAILWRR 145
```

RESULT 31
T00474
hypothetical protein At2g34920 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F19I3.15
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C;Accession: T00474; E84762
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
submitted to the EMBL Data Library, April 1998
A;Description: Arabidopsis thaliana chromosome II BAC F19I3 genomic sequence.
A;Reference number: Z14160
A;Accession: T00474
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-785 <ROU>
A;Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033388
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84762
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <STO>
A;Cross-references: GB:AE002093; NID:g3033388; PIDN:AAC12832.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g34920; F19I3.15
A;Map position: 2
A;Introns: 33/2; 49/3; 95/1; 146/2; 376/3; 415/2; 607/2; 695/3; 745/2

  Query Match 14.5%; Score 61.5; DB 2; Length 785;
  Best Local Similarity 29.6%; Pred. No. 84;
  Matches 24; Conservative 10; Mismatches 44; Indels 3; Gaps 2;

```
Qy          2 GRSGCSSQS--ISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCL- 58
              | |   |||  ||   |: |  |:|    |:     |: :   :| :| |       :
Db        236 GNSAIHSQSIEISSEASVQEIHLLAPSIDGESESENESKSPDQTVEIESGTLNSVSDIIR 295

Qy         59 RLGTHGSPTASSQSSATNMAI 79
              ||      |||:   | :| |
Db        296 RLSNEQKLTASNNGGAVDMPI 316
```

RESULT 32
E72536

probable oligopeptide transport ATP-binding protein APE1578 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: E72536
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: E72536
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-324 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80578.1; PID:g5105265
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1578
C;Superfamily: inner membrane protein malK; ATP-binding cassette homology
F;25-231/Domain: ATP-binding cassette homology <ABC>

  Query Match 14.4%; Score 61; DB 2; Length 324;
  Best Local Similarity 28.9%; Pred. No. 33;
  Matches 22; Conservative 10; Mismatches 22; Indels 22; Gaps 3;

```
Qy          9 QSISPMRSISENSLVAMDFSGQKSRVI--------------ENPTEALSVAVE------- 47
              :|:   |||::     :   ||| ||:               : || || | |:
Db        139 ESVGLHRSIADRYPHELS-GGQKQRVVIAMALALEPDIVIADEPTTALDVVVQAQILNLL 197

Qy         48 EGLAWRKKGCLRLGTH 63
              : ||| |   : | ||
Db        198 KKLAWEKNLSIILITH 213
```

RESULT 33
TNBE12
74K alpha trans-inducing protein - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: C27342
R;Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A;Title: The complete DNA sequence of varicella-zoster virus.
A;Reference number: A27345; MUID:86306657; PMID:3018124
A;Accession: C27342
A;Molecule type: DNA
A;Residues: 1-661 <DAV>
A;Cross-references: EMBL:X04370; NID:g59989; PIDN:CAA27895.1; PID:g60001
C;Genetics:
A;Gene: 12
C;Superfamily: herpesvirus 77K alpha trans-inducing protein
C;Keywords: trans-inducing protein; transcription regulation

Query Match 14.4%; Score 61; DB 1; Length 661;
Best Local Similarity 38.6%; Pred. No. 78;
Matches 22; Conservative 6; Mismatches 15; Indels 14; Gaps 4;

```
Qy          12 SPMRSISENSLVAMDFSGQK-SRVIENPTEALSVAVEEGLAWRKKGCLRLG-THGSP 66
               :|: ||      | | : |: |||:  |   | | |      ||  || :| | |||
Db         506 APLNSI------APDTNRQRTSRVLVRPDTGLDVTV------RKNHCLDIGHTDGSP 550
```

RESULT 34
S44876
ZC21.4 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C;Accession: S44876
R;Du, Z.; Waterston, R.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid ZC21.
A;Reference number: S44649
A;Accession: S44876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-733 <DUZ>
A;Cross-references: EMBL:L16685; NID:g289729; PID:g289735
C;Genetics:
A;Introns: 269/3; 551/3; 600/2; 670/3

Query Match 14.4%; Score 61; DB 2; Length 733;
Best Local Similarity 33.9%; Pred. No. 88;
Matches 19; Conservative 6; Mismatches 27; Indels 4; Gaps 2;

```
Qy          32 SRVIEN---PTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
               || |||   || | :   : |  |:|    : |: |  : ||  |    |    ||
Db          18 SRDIENGEAPT-ATATTPKSGRKWKKSKAAKQGSGGGSSGSSSGSQQQGAAGAPQP 72
```

RESULT 35
W2WLE
E2 protein - human papillomavirus type 1a
C;Species: human papillomavirus type 1a
C;Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 16-Feb-1997
C;Accession: A03665
R;Danos, O.; Katinka, M.; Yaniv, M.
EMBO J. 1, 231-236, 1982
A;Title: Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae.
A;Reference number: A90970; MUID:84182467; PMID:6325156
A;Accession: A03665
A;Molecule type: DNA
A;Residues: 1-322 <DAN>
C;Superfamily: papillomavirus E2 protein
C;Keywords: DNA binding; early protein; transcription regulation

Query Match 14.3%; Score 60.5; DB 1; Length 322;
Best Local Similarity 30.2%; Pred. No. 38;
Matches 19; Conservative 15; Mismatches 22; Indels 7; Gaps 3;

```
Qy          25 MDFSGQKSRVIENPTEALSVAVEEGLAW---RKKGCLRLGTHGSPT-ASSQ---SSATNM 77
               |:   | |::||:  :  ::  :| : :   || | :|:|    |: ||||    :|  |
Db          16 MNLYEQDSKLIEDQIKQWNLIRQEQVLFHFARKNGVMRIGLQAVPSLASSQEKAKTAIEM 75

Qy          78 AIH 80
                :|
Db          76 VLH 78
```

RESULT 36
S53975
probable membrane protein YMR305c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9952.07c
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S53975
R;Connor, R.; Churcher, C.M.
submitted to the EMBL Data Library, April 1995
A;Reference number: S53969
A;Accession: S53975
A;Molecule type: DNA
A;Residues: 1-389 <CON>
A;Cross-references: EMBL:Z49212; NID:g798940; PID:g798947; GSPDB:GN00013;
MIPS:YMR305c
C;Genetics:
A;Gene: SGD:SCW10; MIPS:YMR305c
A;Cross-references: SGD:S0004921
A;Map position: 13R
C;Keywords: transmembrane protein
F;6-22/Domain: transmembrane #status predicted <TMM>

  Query Match   14.3%; Score 60.5; DB 2; Length 389;
  Best Local Similarity   28.6%; Pred. No. 47;
  Matches   22; Conservative   15; Mismatches   33; Indels   7; Gaps   2;

```
Qy           4 SGCSSQSISPMRSISENSLVAMDFS----GQKSRVIENPTEALSVAVEEGLAWRKKGCLR 59
               || | ::| |   ::||::||   |     | :     || : :|   :      :
Db          45 SGNSGETIVP---VNENAVVATTSSTAVASQATTSTLEPTTSANVVTSQQQTSTLQSSEA 101

Qy          60 LGTHGSPTASSQSSATN 76
                 | || |:|| ||:::
Db         102 ASTVGSSTSSSPSSSSS 118
```

RESULT 37
AE1323
3-isopropylmalate dehydratase (large chain) homolog leuC [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AE1323
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-462 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00067.1; PID:g16411442; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: leuC
C;Superfamily: aconitate hydratase

  Query Match 14.3%; Score 60.5; DB 2; Length 462;
  Best Local Similarity 23.8%; Pred. No. 58;
  Matches 15; Conservative 15; Mismatches 20; Indels 13; Gaps 3;

```
Qy          6 CSSQSISPMRSIS--------ENSLVAMDFSGQKSRVIENPTEAL---SVAVEEGLAWRK 54
              |::  :| :   :         :|:: |:   |  || : |  |::    : :| |  ||:
Db        337 CTNARLSDLEEAARIVKGNKVKNNIRALVVPG--SRQVRNAAESIGLDKIFIEAGFEWRE 394

Qy         55 KGC 57
               ||
Db        395 PGC 397
```

RESULT 38
AD0107
hypothetical protein YPO0873 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AD0107
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; Sebaihia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; Baker, S.; Basham, D.; Bentley, S.D.; Brooks, K.; Cerdeno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Feltwell, T.; Hamlin, N.; Holroyd, S.; Jagels, K.; Leather, S.; Karlyshev, A.V.; Moule, S.; Oyston, P.C.F.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.G.
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AD0107
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89719.1; PID:g15978946; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0873

  Query Match 14.3%; Score 60.5; DB 2; Length 512;
  Best Local Similarity 29.9%; Pred. No. 65;
  Matches 20; Conservative 8; Mismatches 30; Indels 9; Gaps 3;

```
Qy          6 CSSQSISPMRSISENSLVAMDFSGQKS------RVIENPTEALSVAVEEGLAWRKKGCLR 59
              | :: |  | :     |:    ||| |         |:| :   |  :  |  |:|||
Db         77 CKARFIPSMMN-DAYELIGSPTSGQSSIAPSFTETSESPPDVTPVFAKSCL--REKGCTD 133

Qy         60 LGTHGSP 66
               || | |
Db        134 AGTEGEP 140
```

RESULT 39
A32608
thyroid hormone receptor-related protein Rev-ErbA-alpha - human
N;Alternate names: erbA-related protein 1; thyroid hormone-binding protein homolog ear-1; transcription factor ear-1
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 23-Mar-1995 #text_change 20-Sep-1999
C;Accession: A32286; A32608; S06164
R;Miyajima, N.; Horiuchi, R.; Shibuya, Y.; Fukushige, S.; Matsubara, K.; Toyoshima, K.; Yamamoto, T.
Cell 57, 31-39, 1989
A;Title: Two erbA homologs encoding proteins with different T-3 binding capacities are transcribed from opposite DNA strands of the same genetic locus.
A;Reference number: A32286; MUID:89195219; PMID:2539258
A;Accession: A32286
A;Molecule type: mRNA
A;Residues: 1-614 <MIY>
A;Cross-references: GB:M24898; NID:g537519; PIDN:AAA52335.1; PID:g537520
R;Lazar, M.A.; Jones, K.E.; Chin, W.W.
DNA Cell Biol. 9, 77-83, 1990
A;Title: Isolation of a cDNA encoding human Rev-ErbA-alpha: transcription from the noncoding DNA strand of a thyroid hormone receptor gene results in a related protein that does not bind thyroid hormone.
A;Reference number: A32608; MUID:90262650; PMID:1971514
A;Accession: A32608
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-146,'L',148-563,'Q',565-614 <LAZ>
R;Miyajima, N.; Kadowaki, Y.; Fukushige, S.; Shimizu, S.; Semba, K.; Yamanashi, Y.; Matsubara, K.; Toyoshima, K.; Yamamoto, T.
Nucleic Acids Res. 16, 11057-11074, 1988
A;Title: Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other.
A;Reference number: S02709; MUID:89083547; PMID:2905047
A;Accession: S06164
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 132-198 <MI2>
C;Comment: Reference A32608 reports that this protein does not bind T-3, while reference A32286 describes low but appreciable binding.
C;Genetics:
A;Gene: ear-1
C;Superfamily: unassigned erbA-related proteins; erbA transforming protein homology
C;Keywords: DNA binding; zinc finger

F;130-548/Domain: erbA transforming protein homology <ERBA>
F;132-198/Domain: DNA binding #status predicted <DNA>
F;132-152/Region: zinc finger
F;169-193/Region: zinc finger

  Query Match 14.3%; Score 60.5; DB 2; Length 614;
  Best Local Similarity 27.1%; Pred. No. 81;
  Matches 23; Conservative 13; Mismatches 44; Indels 5; Gaps 2;

```
Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCL--R 59
              | || |    ||    |:||  :     |       |:   |:  :     |  | :
Db         18 GSSGSSPSRTSPESLYSDNSNGSFQSLTQGCPTYFPPSPTGSLTQDPA---RSFGSIPPS 74

Qy         60 LGTHGSPTASSQSSATNMAIHRSQP 84
              |   |||::|| ||::: : :   |
Db         75 LSDDGSPSSSSSSSSSSSFYNGSP 99
```

RESULT 40
T47449
hypothetical protein T14D3.30 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47449
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.;
Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24467
A;Accession: T47449
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-716 <JOR>
A;Cross-references: EMBL:AL138649
A;Experimental source: cultivar Columbia; BAC clone T14D3
C;Genetics:
A;Map position: 3
A;Introns: 50/3; 150/2; 177/3; 308/3; 548/3; 589/3
A;Note: T14D3.30

  Query Match 14.3%; Score 60.5; DB 2; Length 716;
  Best Local Similarity 28.6%; Pred. No. 98;
  Matches 30; Conservative 14; Mismatches 32; Indels 29; Gaps 7;

```
Qy          2 GRSGCSSQSISPMRS--------ISENSLVAM--DFSGQKSRVIENPT---------EAL 42
              | |||   ::| :          :| :|:  |   |: :|    :||          | |
Db        164 GTSGCGKSTLSALLGSRLGITTVVSTDSIRHMMRSFADEK----QNPLLWASTYHAGEYL 219

Qy         43 S-VAVEEGLAWRK----KGCLRLGTHGSPT-ASSQSSATNMAIHR 81
                |||  |  | ||    ||   : ::   | | | || | :  |:
Db        220 DPVAVAESKAKRKAKKLKGSRGVNSNAQKTDAGSNSSTTELLSHK 264
```

Search completed: January 13, 2004, 16:24:13
Job time : 21.5197 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:22:54 ; Search time 36.378 Seconds
(without alignments)
465.304 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

1	423	100.0	540	15	US-10-242-332-2	Sequence 2, Appli
2	423	100.0	540	16	US-10-323-001-2	Sequence 2, Appli
3	191	45.2	375	12	US-10-094-749-3245	Sequence 3245, Ap
4	191	45.2	535	15	US-10-242-332-3	Sequence 3, Appli
5	191	45.2	535	16	US-10-323-001-3	Sequence 3, Appli
6	186	44.0	621	15	US-10-242-332-4	Sequence 4, Appli
7	186	44.0	621	16	US-10-323-001-4	Sequence 4, Appli
8	179	42.3	532	15	US-10-097-340-125	Sequence 125, App
9	179	42.3	532	15	US-10-233-098-2	Sequence 2, Appli
10	68.5	16.2	537	14	US-10-037-667-1	Sequence 1, Appli
11	66.5	15.7	564	12	US-10-369-493-19159	Sequence 19159, A
12	65.5	15.5	541	15	US-10-230-026-44	Sequence 44, Appl
13	65	15.4	156	9	US-09-925-301-1154	Sequence 1154, Ap
14	65	15.4	653	14	US-10-023-437-67	Sequence 67, Appl
15	64.5	15.2	196	12	US-10-287-274-379	Sequence 379, App
16	63	14.9	556	12	US-10-369-493-12607	Sequence 12607, A
17	63	14.9	754	12	US-10-369-493-8297	Sequence 8297, Ap
18	62.5	14.8	663	12	US-10-104-047-3473	Sequence 3473, Ap
19	62	14.7	431	10	US-09-764-864-820	Sequence 820, App
20	62	14.7	1753	15	US-10-146-473-44	Sequence 44, Appl
21	62	14.7	2344	9	US-09-815-242-12713	Sequence 12713, A
22	61.5	14.5	1047	9	US-09-866-562-57	Sequence 57, Appl
23	61.5	14.5	1616	12	US-10-205-219-119	Sequence 119, App
24	61	14.4	99	9	US-09-864-761-36007	Sequence 36007, A
25	61	14.4	128	12	US-10-029-386-33561	Sequence 33561, A
26	61	14.4	465	15	US-10-156-761-9029	Sequence 9029, Ap
27	60.5	14.3	489	12	US-10-369-493-4345	Sequence 4345, Ap
28	60.5	14.3	497	12	US-10-369-493-7100	Sequence 7100, Ap
29	60.5	14.3	674	15	US-10-090-455-4	Sequence 4, Appli
30	60.5	14.3	2861	12	US-10-374-979-108	Sequence 108, App
31	60.5	14.3	2861	12	US-10-331-496A-89	Sequence 89, Appl
32	60.5	14.3	3038	12	US-09-863-776-62	Sequence 62, Appl
33	60	14.2	310	12	US-10-306-292-27	Sequence 27, Appl
34	59.5	14.1	1346	11	US-09-793-708-4	Sequence 4, Appli
35	59.5	14.1	1346	12	US-10-201-365-5	Sequence 5, Appli
36	59.5	14.1	1346	12	US-10-160-539-4	Sequence 4, Appli
37	59	13.9	246	9	US-09-815-242-13184	Sequence 13184, A
38	59	13.9	638	14	US-10-072-621-10	Sequence 10, Appl
39	59	13.9	1024	15	US-10-211-962-85	Sequence 85, Appl
40	58.5	13.8	189	12	US-10-104-047-3196	Sequence 3196, Ap
41	58.5	13.8	573	9	US-09-815-242-11257	Sequence 11257, A
42	58.5	13.8	602	12	US-10-094-749-3150	Sequence 3150, Ap
43	58.5	13.8	652	10	US-09-992-647-1	Sequence 1, Appli
44	58.5	13.8	652	15	US-10-225-567A-653	Sequence 653, App
45	58.5	13.8	661	9	US-09-764-853-679	Sequence 679, App

ALIGNMENTS

RESULT 1
US-10-242-332-2
; Sequence 2, Application US/10242332
; Publication No. US20030044834A1
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger John

```
;  APPLICANT:  Sutherland, Robert Lyndsay
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001710
;  CURRENT APPLICATION NUMBER: US/10/242,332
;  CURRENT FILING DATE:  2002-09-11
;  PRIOR APPLICATION NUMBER: US 08/945,771
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: PCT/AU96/00258
;  PRIOR FILING DATE: 1996-05-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-242-332-2

  Query Match             100.0%;  Score 423;  DB 15;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-43;
  Matches   84;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 414

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db        415 GTHGSPTASSQSSATNMAIHRSQP 438


RESULT 2
US-10-323-001-2
; Sequence 2, Application US/10323001
; Publication No. US20030129639A1
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger John
;  APPLICANT:  Sutherland, Robert Lyndsay
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001710
;  CURRENT APPLICATION NUMBER: US/10/323,001
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: US/10/242,332
;  PRIOR FILING DATE: 2002-09-11
;  PRIOR APPLICATION NUMBER: US 08/945,771
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: PCT/AU96/00258
;  PRIOR FILING DATE: 1996-05-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-323-001-2

  Query Match             100.0%;  Score 423;  DB 16;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-43;
```

```
  Matches   84;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 414

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db        415 GTHGSPTASSQSSATNMAIHRSQP 438


RESULT 3
US-10-094-749-3245
; Sequence 3245, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3245
;   LENGTH: 375
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-749-3245

  Query Match             45.2%;  Score 191;  DB 12;  Length 375;
  Best Local Similarity   59.7%;  Pred. No. 7.2e-15;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        206 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 262
```

```
Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         263 S-LSAAIHRTQP 273


RESULT 4
US-10-242-332-3
; Sequence 3, Application US/10242332
; Publication No. US20030044834A1
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger John
;  APPLICANT:  Sutherland, Robert Lyndsay
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001710
;  CURRENT APPLICATION NUMBER: US/10/242,332
;  CURRENT FILING DATE:  2002-09-11
;  PRIOR APPLICATION NUMBER: US 08/945,771
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: PCT/AU96/00258
;  PRIOR FILING DATE: 1996-05-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 535
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-242-332-3

  Query Match             45.2%;  Score 191;  DB 15;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 1.1e-14;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         423 S-LSAAIHRTQP 433


RESULT 5
US-10-323-001-3
; Sequence 3, Application US/10323001
; Publication No. US20030129639A1
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger John
;  APPLICANT:  Sutherland, Robert Lyndsay
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001710
;  CURRENT APPLICATION NUMBER: US/10/323,001
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: US/10/242,332
;  PRIOR FILING DATE: 2002-09-11
;  PRIOR APPLICATION NUMBER: US 08/945,771
;  PRIOR FILING DATE: 1998-04-22
```

```
;  PRIOR APPLICATION NUMBER: PCT/AU96/00258
;  PRIOR FILING DATE: 1996-05-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 535
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-323-001-3

  Query Match             45.2%;  Score 191;  DB 16;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 1.1e-14;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:||:|:|:|||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db         366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422

Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         423 S-LSAAIHRTQP 433


RESULT 6
US-10-242-332-4
; Sequence 4, Application US/10242332
; Publication No. US20030044834A1
; GENERAL INFORMATION:
;  APPLICANT: Daly, Roger John
;  APPLICANT:  Sutherland, Robert Lyndsay
;  TITLE OF INVENTION: GDU, A novel signalling protein
;  FILE REFERENCE: 273402001710
;  CURRENT APPLICATION NUMBER: US/10/242,332
;  CURRENT FILING DATE:  2002-09-11
;  PRIOR APPLICATION NUMBER: US 08/945,771
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: PCT/AU96/00258
;  PRIOR FILING DATE: 1996-05-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 621
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-242-332-4

  Query Match             44.0%;  Score 186;  DB 15;  Length 621;
  Best Local Similarity   54.1%;  Pred. No. 5.6e-14;
  Matches   46;  Conservative    6;  Mismatches   23;  Indels   10;  Gaps    3;

Qy           3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
               | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db         440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy          63 HGSPTASSQS----SATNMAIHRSQ 83
                     ||||    |  |  |||:|
Db         498 ----ILSSQSPLHPSTLNAVIHRTQ 518
```

RESULT 7
US-10-323-001-4
; Sequence 4, Application US/10323001
; Publication No. US20030129639A1
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger John
; APPLICANT: Sutherland, Robert Lyndsay
; TITLE OF INVENTION: GDU, A novel signalling protein
; FILE REFERENCE: 273402001710
; CURRENT APPLICATION NUMBER: US/10/323,001
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US/10/242,332
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 08/945,771
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT/AU96/00258
; PRIOR FILING DATE: 1996-05-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-323-001-4

Query Match 44.0%; Score 186; DB 16; Length 621;
Best Local Similarity 54.1%; Pred. No. 5.6e-14;
Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

```
Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||||  |||:|| || | |:||| |||  |  |:
Db        440 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWR-NGSTRMN- 497

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                    ||||    |  |  |||:|
Db        498 ----ILSSQSPLHPSTLNAVIHRTQ 518
```

RESULT 8
US-10-097-340-125
; Sequence 125, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.

; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The
Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 125
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-125

Query Match 42.3%; Score 179; DB 15; Length 532;
Best Local Similarity 59.2%; Pred. No. 3.3e-13;
Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

```
Qy          13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
               |:|| |:|:|||||||   |||||| ||||||:||  |||||   ||     |  :| :
Db         363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 419

Qy          73 SATNMAIHRSQ 83
               |  : ||||:|
Db         420 S-LSAAIHRTQ 429
```

RESULT 9
US-10-233-098-2
; Sequence 2, Application US/10233098
; Publication No. US20030109440A1
; GENERAL INFORMATION:
; APPLICANT: Chu, Peter
; APPLICANT: Li, Congfen
; APPLICANT: Liao, X. Charlene
; APPLICANT: Masuda, Esteban
; APPLICANT: Pardo, Jorge
; APPLICANT: Zhao, Haoran
; APPLICANT: Rigel Pharmaceuticals, Incorporated

```
;  TITLE OF INVENTION: GRB7:  No. US20030109440A1el Regulator of Lymphocytic
Signaling
;  FILE REFERENCE: 021044-004500
;  CURRENT APPLICATION NUMBER: US/10/233,098
;  CURRENT FILING DATE:  2002-08-30
;  PRIOR APPLICATION NUMBER: US 60/327,212
;  PRIOR FILING DATE: 2001-10-03
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 532
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human wild-type growth factor receptor-bound 7
;   OTHER INFORMATION: (GRB7)
US-10-233-098-2

  Query Match             42.3%;  Score 179;  DB 15;  Length 532;
  Best Local Similarity   59.2%;  Pred. No. 3.3e-13;
  Matches   42;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:|| |:|:||||||||   |||||| ||||||:||  |||||   ||      |  :| :
Db        363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 419

Qy         73 SATNMAIHRSQ 83
              |  : ||||:|
Db        420 S-LSAAIHRTQ 429


RESULT 10
US-10-037-667-1
; Sequence 1, Application US/10037667
; Publication No. US20020177145A1
; GENERAL INFORMATION:
;  APPLICANT: Morgan, Bruce A.
;  TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY
;  TITLE OF INVENTION:  DAEDALOS
;  FILE REFERENCE: 10287-044001
;  CURRENT APPLICATION NUMBER: US/10/037,667
;  CURRENT FILING DATE:  2002-07-23
;  PRIOR APPLICATION NUMBER: 60/243,110
;  PRIOR FILING DATE: 2000-10-25
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 537
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-037-667-1

  Query Match             16.2%;  Score 68.5;  DB 14;  Length 537;
  Best Local Similarity   38.0%;  Pred. No. 11;
  Matches   19;  Conservative    9;  Mismatches   17;  Indels    5;  Gaps    1;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 51
```

```
              :||  || ||:| ||:    :| ::|  :  | | |     || ||: |:
Db          31 NSQHSSPSRSLSANSIKVEMYSDEESSRLLGPDERLLDKDDSVIVEDSLS 80
```

RESULT 11
US-10-369-493-19159
; Sequence 19159, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19159
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19159

Query Match 15.7%; Score 66.5; DB 12; Length 564;
Best Local Similarity 35.3%; Pred. No. 21;
Matches 24; Conservative 11; Mismatches 24; Indels 9; Gaps 3;

```
Qy           4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPT-EALSVAVEEGLAWRKKGCLRLGT 62
               :| :|: |||        |||||  | :  |: ||: | |:|  ||   :::   | | |
Db         203 AGRASEQISP------GDLVAMD--GIRGVVLVNPSDEQLAVFREEQRRYQESERLALAT 254

Qy          63 HGSPTASS 70
                 |  |:
Db         255 KDLPAVST 262
```

RESULT 12
US-10-230-026-44
; Sequence 44, Application US/10230026
; Publication No. US20030124695A1
; GENERAL INFORMATION:
; APPLICANT: MICHAEL G. BRAMUCCI
; APPLICANT: PATRICIA C. BRZOSTOWICZ
; APPLICANT: KRISTY N. KOSTICHKA
; APPLICANT: VASANTHA NAGARAJAN
; APPLICANT: PIERRE E. ROUVIERE
; APPLICANT: STUART M. THOMAS
; TITLE OF INVENTION: GENES ENCODING BAEYER-VILLIGER MONOOXYGENASES
; FILE REFERENCE: CL1789 US NA
; CURRENT APPLICATION NUMBER: US/10/230,026
; CURRENT FILING DATE: 2002-08-28

```
;  PRIOR APPLICATION NUMBER: 60/315,546
;  PRIOR FILING DATE: 2001-08-29
;  NUMBER OF SEQ ID NOS: 113
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 44
;   LENGTH: 541
;   TYPE: PRT
;   ORGANISM: Rhodococcus erythropolis AN12
US-10-230-026-44

  Query Match             15.5%;  Score 65.5;  DB 15;  Length 541;
  Best Local Similarity   26.0%;  Pred. No. 26;
  Matches   25;  Conservative   15;  Mismatches   35;  Indels   21;  Gaps    4;

Qy          2 GRSGCSSQSISPMRSISEN-------SLVAMDFSGQKSRVIENPTEALSVAVEEGL---- 50
              |:   : : |  :::  ||       | ||  |        |:   |:||: || |
Db        219 GKRAVTDEQIDAVKADYENIWTQVKRSSVAFGFE-------ESTVPAMSVSAEERLRVYE 271

Qy         51 -AWRKKGCLR--LGTHGSPTASSQSSATNMAIHRSQ 83
               || : |  |   || |      ::: |  :  ||:
Db        272 EAWEQGGGFRFMFGTFGDIATDEEANETAASFIRSK 307


RESULT 13
US-09-925-301-1154
; Sequence 1154, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1154
;   LENGTH: 156
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-301-1154

  Query Match             15.4%;  Score 65;  DB 9;  Length 156;
  Best Local Similarity   28.1%;  Pred. No. 6.1;
  Matches   27;  Conservative   12;  Mismatches   35;  Indels   22;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKS----------------RVIENPTEALSVAV 46
              ||  || | |   | | :|  :   ||  |                 : : :|| |   ||
Db          6 RSSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 65

Qy         47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
               ||       | ||:      :  | ::|| ||:::
Db         66 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 101
```

RESULT 14
US-10-023-437-67
; Sequence 67, Application US/10023437
; Publication No. US20020183272A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: STEMKE-HALE, KATHERINE
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: KALTENBOECK, BERNHARD
; TITLE OF INVENTION: METHODS AND compositions for Vaccination COMPRISING
NUCLEIC ACID
; TITLE OF INVENTION: AND/OR POLYPEPTIDE SEQUENCES OF CHLAMYDIA
; FILE REFERENCE: UTSD:736US
; CURRENT APPLICATION NUMBER: US/10/023,437
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/225,839
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Chlamydia psittaci
US-10-023-437-67

Query Match 15.4%; Score 65; DB 14; Length 653;
Best Local Similarity 31.3%; Pred. No. 38;
Matches 26; Conservative 12; Mismatches 35; Indels 10; Gaps 3;

```
Qy          2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
              || |      ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db        358 GRKG------SPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG 411

Qy         62 THGSPTA---SSQSSATNMAIHR 81
                |  |     |::|    : ||
Db        412 -FGRKTGIELPSEASGLVPSPHR 433
```

RESULT 15
US-10-287-274-379
; Sequence 379, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE
THERETO
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 379
;   LENGTH: 196
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-10-287-274-379

  Query Match             15.2%;  Score 64.5;  DB 12;  Length 196;
  Best Local Similarity   28.0%;  Pred. No. 9.4;
  Matches   21;  Conservative   15;  Mismatches   24;  Indels   15;  Gaps    4;

Qy          17 ISENSLV-AMDFSGQKSR------VIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTAS 69
               | |||:| |  |   |:       :: :| :|:    |: |||:|:     |||
Db         109 IGENSIVGASAFVKAKAEMPANYLIVGSPAKAIRELSEQELAWKKQ-----GTHEYQVLV 163

Qy          70 SQSSATNMAIHRSQP 84
               ::   |   :|: :|
Db         164 TRCKQT---LHQVEP 175


RESULT 16
US-10-369-493-12607
; Sequence 12607, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12607
;   LENGTH: 556
;   TYPE: PRT
;   ORGANISM: Aspergillus nidulans
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(556)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12607

  Query Match             14.9%;  Score 63;  DB 12;  Length 556;
  Best Local Similarity   28.6%;  Pred. No. 55;
  Matches   26;  Conservative   19;  Mismatches   32;  Indels   14;  Gaps    4;

Qy           3 RSGCSSQSISPMRSISEN--SLVAMDFSGQ----KSRVIENP------TEALSVAVEEGL 50
               | | :: |:  :    ||   |:  :|||      | :  ||      :||:: |: | :
Db         421 RDGMNTTSLEYILCQQENHSPLILSEFSGTAGALSSAIHINPWDTIGVSEAINKALTESV 480
```

```
Qy          51 AWRKKGCLRLGTHGSPTASSQSSATNMAIHR 81
               | :|:  |:|  |   | :: |: :|  | |
Db         481 ADKKEQHLKLYKH--VTTNTVSAWSNQFISR 509
```

RESULT 17
US-10-369-493-8297
; Sequence 8297, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8297
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-369-493-8297

Query Match 14.9%; Score 63; DB 12; Length 754;
Best Local Similarity 30.0%; Pred. No. 81;
Matches 21; Conservative 14; Mismatches 33; Indels 2; Gaps 2;

```
Qy           1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK-GCLR 59
               :| :|  || : |   |: :: ||::  |::   :    |   : :|| |  ||  |  |
Db         254 KGTNG-KSQGVVPFLKIANDTAVAVNQGGKRKGAVCAYLETWHIDIEEFLDLRKNTGDER 312

Qy          60 LGTHGSPTAS 69
                 ||   ||:
Db         313 RRTHDMNTAN 322
```

RESULT 18
US-10-104-047-3473
; Sequence 3473, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 3473
;   LENGTH: 663
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3473

  Query Match             14.8%;  Score 62.5;  DB 12;  Length 663;
  Best Local Similarity   30.1%;  Pred. No. 79;
  Matches   25;  Conservative   11;  Mismatches   22;  Indels   25;  Gaps    4;

Qy          8 SQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSV-------------------AVE 47
              |:|:||    |  :|  :    | :|  |:  :|| |                    :|:
Db        221 SESMSPGDPCSSRALQVLSIGSQWARA-EDALQALKVGEKPPTWEVTLGASVRASSGSVQ 279

Qy         48 EGLAWRKKGCLRLGTHGSPTASS 70
              | |  |  |   ||| |:|:|||
Db        280 EDL--RSTGA--LGTTGNPSASS 298


RESULT 19
US-09-764-864-820
; Sequence 820, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 820
;   LENGTH: 431
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-820

  Query Match             14.7%;  Score 62;  DB 10;  Length 431;
  Best Local Similarity   32.9%;  Pred. No. 52;
  Matches   24;  Conservative    9;  Mismatches   28;  Indels   12;  Gaps    4;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRKKGCLRLG 61
              |: | |  :  :| | |  |  :  | |  || :::|      :||   ||  ||    |
Db        315 RADCLSTGMELLRRIQERLLAILQHSAQDFRVGLQSP------SVE---AWEAKGPNMPG 365

Qy         62 THGSPTASSQSSA 74
              :   | |||   |
Db        366 S--QPQASSGPEA 376


RESULT 20
US-10-146-473-44
; Sequence 44, Application US/10146473
; Publication No. US20030108888A1
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew
```

```
;  APPLICANT:  Gout, Ivan
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Gure, Ali
;  APPLICANT:  Chen, Yao-Tseng
;  APPLICANT:  Old, Lloyd
;  TITLE OF INVENTION: Breast Cancer Antigens
;  FILE REFERENCE: L00461/70130(JRV)
;  CURRENT APPLICATION NUMBER: US/10/146,473
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: US 60/291,150
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 82
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
;   LENGTH: 1753
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-146-473-44

  Query Match           14.7%;  Score 62;  DB 15;  Length 1753;
  Best Local Similarity  32.9%;  Pred. No. 3.2e+02;
  Matches   24;  Conservative    9;  Mismatches   28;  Indels   12;  Gaps    4;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRKKGCLRLG 61
              |: | |  :  :| | |  |  :  | |  || :::|      :||   ||  ||    |
Db       1637 RADCLSTGMELLRRIQERLLAILQHSAQDFRVGLQSP------SVE---AWEAKGPNMPG 1687

Qy         62 THGSPTASSQSSA 74
              :   | |||   |
Db       1688 S--QPQASSGPEA 1698


RESULT 21
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
```

```
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             14.7%;  Score 62;  DB 9;  Length 2344;
  Best Local Similarity   30.3%;  Pred. No. 4.6e+02;
  Matches   23;  Conservative   21;  Mismatches   18;  Indels   14;  Gaps    4;

Qy          8 SQSISPMRSISENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRKKGCLRLGTHGSP 66
              | |:|   |||::: :::  || :| |  |: ::: |::  | |:          | ||
Db       2014 STSLSTSDSISDSTSISI--SGSQSAVESESTSDSTSISDSESLS----------TSGS- 2060

Qy         67 TASSQSSATNMAIHRS 82
              |:|| |::|: ::  |
Db       2061 TSSSTSTSTSESLSTS 2076


RESULT 22
US-09-866-562-57
; Sequence 57, Application US/09866562
; Patent No. US20020009758A1
; GENERAL INFORMATION:
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Klee, Jennifer
;  APPLICANT:  Switzer, Anne
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER.
;  FILE REFERENCE: 210121.502
;  CURRENT APPLICATION NUMBER: US/09/866,562
;  CURRENT FILING DATE:  2001-05-25
;  NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 57
;   LENGTH: 1047
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-866-562-57

  Query Match             14.5%;  Score 61.5;  DB 9;  Length 1047;
  Best Local Similarity   32.8%;  Pred. No. 1.9e+02;
  Matches   21;  Conservative    8;  Mismatches   28;  Indels    7;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              || |: :   || | |  | |  |:  :|| | |      ::|  |    | || |  |
Db         20 PMDSLIQELSVAYDCSMAKKRTAED--QALGVPVN-----KRKSLLMKPRHYSPKADCQE 72
```

```
Qy          73 SATN 76
                : :
Db          73 DRSD 76
```

RESULT 23
US-10-205-219-119
; Sequence 119, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Phosphacan
US-10-205-219-119

Query Match 14.5%; Score 61.5; DB 12; Length 1616;
Best Local Similarity 35.4%; Pred. No. 3.3e+02;
Matches 28; Conservative 7; Mismatches 27; Indels 17; Gaps 5;

```
Qy           7 SSQSISPMRSISENSLV---AMDFSGQKSRVIE-----NPTEALSVAVEEGLAWRKKGCL 58
               :| |:| : |:   |||      |  : ||| |      ||   | |   |  | | |
Db        1096 TSVSVSSINSVFTESLVYPITKVFDQEISRVPEIIFPVKPTHTASQA--SGDTWLKPG-- 1151

Qy          59 RLGTHGSP----TASSQSS 73
                | |:  |    ||||: |
Db        1152 -LSTNSEPALSDTASSEVS 1169
```

RESULT 24
US-09-864-761-36007
; Sequence 36007, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36007
;   LENGTH: 99
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078639.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 34
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 8.3
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.4
;   OTHER INFORMATION: EST_HUMAN HIT: H18350.1, EVALUE 9.90e-01
US-09-864-761-36007

```
  Query Match             14.4%;  Score 61;  DB 9;  Length 99;
  Best Local Similarity   26.6%;  Pred. No. 10;
  Matches   21;  Conservative   18;  Mismatches   22;  Indels   18;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVA--MDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              :|| | :|:|  :| | |  ::  :  | :||      :   |:| |:  : ||
Db          3 KSGSSRKSVSSSKSTSSNKAMSSRLSMSSRKSL-----SSLKSIASEKSRSSRKS----- 52

Qy         61 GTHGSPTASSQSSATNMAI 79
                     :||:|:::| |:
Db         53 ------VSSSKSTSSNKAM 65


RESULT 25
US-10-029-386-33561
; Sequence 33561, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33561
;   LENGTH: 128
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078639.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: SWISSPROT HIT: Q90508, EVALUE 8.00e-02
US-10-029-386-33561

  Query Match             14.4%;  Score 61;  DB 12;  Length 128;
  Best Local Similarity   26.6%;  Pred. No. 15;
  Matches   21;  Conservative   18;  Mismatches   22;  Indels   18;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVA--MDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              :|| | :|:|  :| | |  ::  :  | :||      :   |:| |:  : ||
Db          3 KSGSSRKSVSSSKSTSSNKAMSSRLSMSSRKSL-----SSLKSIASEKSRSSRKS----- 52

Qy         61 GTHGSPTASSQSSATNMAI 79
                     :||:|:::| |:
Db         53 ------VSSSKSTSSNKAM 65


RESULT 26
US-10-156-761-9029
; Sequence 9029, Application US/10156761
```

; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9029
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9029

Query Match 14.4%; Score 61; DB 15; Length 465;
Best Local Similarity 36.7%; Pred. No. 76;
Matches 18; Conservative 6; Mismatches 23; Indels 2; Gaps 2;

```
Qy          33 RVIENPT-EALSVAVEEGLAWRKK-GCLRLGTHGSPTASSQSSATNMAI 79
               ||:|:|    |  |||:   |: | | :     |:| ||    |   ||
Db          19 RVVEHPAWPVLKDAVEQIRPWQSKDGSIDFEAEGAPDASDAELAVRRAI 67
```

RESULT 27
US-10-369-493-4345
; Sequence 4345, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4345
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-369-493-4345

```
  Query Match             14.3%;  Score 60.5;  DB 12;  Length 489;
  Best Local Similarity   27.9%;  Pred. No. 94;
  Matches   19;  Conservative    8;  Mismatches   24;  Indels   17;  Gaps    2;

Qy          26 DFSGQKSRVIENPT--EALSVAVEEGL---------------AWRKKGCLRLGTHGSPTA 68
               |||  :  :  :|     ||:||:||:                ||:   :|  ||| |
Db         257 DFSRMRRGLHVDPELYRRLSLAVDEGINMYGMTETATAFTCGDWREPADVRQSTHGKPFD 316

Qy          69 SSQSSATN 76
               |     |
Db         317 GSDLRICN 324


RESULT 28
US-10-369-493-7100
; Sequence 7100, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7100
;   LENGTH: 497
;   TYPE: PRT
;   ORGANISM: Burkholderia cepacia
US-10-369-493-7100

  Query Match             14.3%;  Score 60.5;  DB 12;  Length 497;
  Best Local Similarity   27.9%;  Pred. No. 96;
  Matches   19;  Conservative    8;  Mismatches   24;  Indels   17;  Gaps    2;

Qy          26 DFSGQKSRVIENPT--EALSVAVEEGL---------------AWRKKGCLRLGTHGSPTA 68
               |||  :  :  :|     ||:||:||:                ||:   :|  ||| |
Db         261 DFSRMRRGLHVDPELYRRLSLAVDEGINMYGMTETATAFTCGDWREPADVRQSTHGKPFD 320

Qy          69 SSQSSATN 76
               |     |
Db         321 GSDLRICN 328


RESULT 29
US-10-090-455-4
; Sequence 4, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Chen, Hongyun
;  APPLICANT:  Le Bihan, Stephane
;  TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
;  FILE REFERENCE: 100103.406
;  CURRENT APPLICATION NUMBER: US/10/090,455
;  CURRENT FILING DATE:  2002-03-01
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 674
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-090-455-4

  Query Match             14.3%;  Score 60.5;  DB 15;  Length 674;
  Best Local Similarity   25.9%;  Pred. No. 1.4e+02;
  Matches   28;  Conservative   13;  Mismatches   30;  Indels   37;  Gaps    5;

Qy          3 RSGCSS--QSISPMRSISENSLVAMDFSGQKSRVIENPTEA-----------------L 42
              :| | |  : :|     :|  |:    :|   :|  | |||                  |
Db         23 KSVCVSVDEVVSSNMEATETDLL----NGHLKKVDNNLTEAQRFSSLPRRAAVNIEFRDL 78

Qy         43 SVAVEEGLAWRKKG--CLRLGTHG----------SPTASSQSSATNM 77
              | :| ||  |||||   |  |  |            |: : :|:  |:
Db         79 SYSVPEGPWWRKKGYKTLLKGISGKFNSGELVAIMGPSGAGKSTLMNI 126


RESULT 30
US-10-374-979-108
; Sequence 108, Application US/10374979
; Publication No. US20030219793A1
; GENERAL INFORMATION:
;  APPLICANT: John P. Carulli et al.
;  TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3
;  FILE REFERENCE: 032796-021
;  CURRENT APPLICATION NUMBER: US/10/374,979
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: US 09/544,398
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: US 09/543,771
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: US 09/229,319
;  PRIOR FILING DATE: 1999-01-13
;  PRIOR APPLICATION NUMBER: US 60/071,449
;  PRIOR FILING DATE: 1998-01-13
;  PRIOR APPLICATION NUMBER: US 60/105,511
;  PRIOR FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 109
; SEQ ID NO 108
;   LENGTH: 2861
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-374-979-108

  Query Match             14.3%;  Score 60.5;  DB 12;  Length 2861;
  Best Local Similarity   21.6%;  Pred. No. 9.1e+02;
  Matches   21;  Conservative   18;  Mismatches   45;  Indels   13;  Gaps    2;
```

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQK--SRVIENPT----------EALSVAVE 47
              :|  |  :  : |  :| ::||:  |    |  ||::  ||           |:   |:
Db       1800 EGEEGADAVPLPPPMAIQQHSLLQPDSQDDKASSRLLVRPTSSETPSAAELVSAIEELVK 1859

Qy         48 EGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
                :|   :    |   |  :: | : : |  :  | |
Db       1860 SKMALEDRPSSLLVDQGDSSSPSFNPSDNSLLSSSSP 1896
```

RESULT 31
US-10-331-496A-89
; Sequence 89, Application US/10331496A
; Publication No. US20030228305A1
; GENERAL INFORMATION:
; APPLICANT: FRANTZ,GRETCHEN
; APPLICANT: HILLAN,KENNETH J.
; APPLICANT: PHILLIPS,HEIDI S.
; APPLICANT: POLAKIS,PAUL
; APPLICANT: SMITH,VICTORIA
; APPLICANT: SPENCER,SUSAN D.
; APPLICANT: WILLIAMS,P. MICKEY
; APPLICANT: WU,THOMAS D.
; APPLICANT: ZHANG,ZEMIN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: P5014R1-PCT
; CURRENT APPLICATION NUMBER: US/10/331,496A
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 60/345,444
; PRIOR FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: US 60/351,885
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/360,066
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: US 60/362,004
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/366,869
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US 60/366,284
; PRIOR FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 60/368,679
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/404,809
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US 60/405,645
; PRIOR FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 89
; LENGTH: 2861
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-331-496A-89

Query Match 14.3%; Score 60.5; DB 12; Length 2861;
Best Local Similarity 21.6%; Pred. No. 9.1e+02;
Matches 21; Conservative 18; Mismatches 45; Indels 13; Gaps 2;

```
Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQK--SRVIENPT-----------EALSVAVE 47
              :|  |  :  : |  :| ::||:  |    |  ||::  ||           |:   |:
Db       1800 EGEEGADAVPLPPPMAIQQHSLLQPDSQDDKASSRLLVRPTSSETPSAAELVSAIEELVK 1859

Qy         48 EGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
                :|   :    |   |  :: | : : |  :  | |
Db       1860 SKMALEDRPSSLLVDQGDSSSPSFNPSDNSLLSSSSP 1896


RESULT 32
US-09-863-776-62
; Sequence 62, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
;  APPLICANT: Spytek, Kimberly A
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier, Raymond J
;  APPLICANT:  Gangolli, Esha
;  TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding
Same
;  FILE REFERENCE: 21402-020
;  CURRENT APPLICATION NUMBER: US/09/863,776
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: 09/540,763
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/206,679
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,688
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,829
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/207,748
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/207,798
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/208,263
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: 60/208,831
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 60/209,451
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: 60/210,060
;  PRIOR FILING DATE: 2000-06-07
;  PRIOR APPLICATION NUMBER: 60/219,507
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/221,337
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/221,927
;  PRIOR FILING DATE: 2000-07-31
```

```
;  PRIOR APPLICATION NUMBER: 60/263,135
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/263,688
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 60/263,694
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
;   LENGTH: 3038
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-863-776-62

  Query Match             14.3%;  Score 60.5;  DB 12;  Length 3038;
  Best Local Similarity   21.6%;  Pred. No. 9.8e+02;
  Matches   21;  Conservative   18;  Mismatches   45;  Indels   13;  Gaps    2;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQK--SRVIENPT----------EALSVAVE 47
              :|  |  :  : |  :| ::||:  |    |  ||::  ||            |:   |:
Db       1800 EGEEGADAVPLPPPMAIQQHSLLQPDSQDDKASSRLLVRPTSSETPSAAELVSAIEELVK 1859

Qy         48 EGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
                :|   :    |   |  :: | : : |  :  | |
Db       1860 SKMALEDRPSSLLVDQGDSSSPSFNPSDNSLLSSSSP 1896


RESULT 33
US-10-306-292-27
; Sequence 27, Application US/10306292
; Publication No. US20030145347A1
; GENERAL INFORMATION:
;  APPLICANT: Lanahan, Michael B.
;  APPLICANT:  Desai, Nalini M.
;  APPLICANT:  Gasdaska, Pamela Y.
;  TITLE OF INVENTION: GRAIN PROCESSING METHOD AND TRANSGENIC PLANTS USEFUL
;  TITLE OF INVENTION:  THEREIN
;  FILE REFERENCE: A-31383P1
;  CURRENT APPLICATION NUMBER: US/10/306,292
;  CURRENT FILING DATE:  2002-11-27
;  PRIOR APPLICATION NUMBER: US/09/598,747
;  PRIOR FILING DATE: 2000-06-21
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-306-292-27

  Query Match             14.2%;  Score 60;  DB 12;  Length 310;
  Best Local Similarity   36.6%;  Pred. No. 60;
  Matches   15;  Conservative    8;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          6 CSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAV 46
              | :||:    ||   :: |:||| :  ||  : |  |: ||
Db         78 CRAQSLRFGTSIISETVTAVDFSARPFRVASDSTTVLADAV 118
```

RESULT 34
US-09-793-708-4
; Sequence 4, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-4

  Query Match   14.1%; Score 59.5; DB 11; Length 1346;
  Best Local Similarity   34.6%; Pred. No. 4.6e+02;
  Matches   18; Conservative   9; Mismatches   14; Indels   11; Gaps   2;

Qy   13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
              |:| | :|| |:|| : :| | |:|| ||: : : || |
Db   972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

RESULT 35
US-10-201-365-5
; Sequence 5, Application US/10201365
; Publication No. US20030148469A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A
MODULAR

; TITLE OF INVENTION: PKS GENE CLUSTER AS SCAFFOLD
; FILE REFERENCE: 300622002103
; CURRENT APPLICATION NUMBER: US/10/201,365
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-10-201-365-5

  Query Match   14.1%; Score 59.5; DB 12; Length 1346;
  Best Local Similarity 34.6%; Pred. No. 4.6e+02;
  Matches   18; Conservative   9; Mismatches   14; Indels   11; Gaps   2;

Qy   13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
                |:| | :|| |:|| : :| | |:|| ||: : : || |
Db   972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

RESULT 36
US-10-160-539-4
; Sequence 4, Application US/10160539
; Publication No. US20030162262A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/10/160,539
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US/09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-10-160-539-4

  Query Match   14.1%; Score 59.5; DB 12; Length 1346;
  Best Local Similarity 34.6%; Pred. No. 4.6e+02;
  Matches   18; Conservative   9; Mismatches   14; Indels   11; Gaps   2;

```
Qy          13 PMRSISENSLVAMDFSGQKSR----------VIENPTE-ALSVAVEEGLAWR 53
               |:| |  :|| |:||  : :|          | |:||  ||:  : : || |
Db         972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023
```

RESULT 37
US-09-815-242-13184
; Sequence 13184, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13184
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13184

Query Match 13.9%; Score 59; DB 9; Length 246;
Best Local Similarity 34.8%; Pred. No. 59;
Matches 16; Conservative 7; Mismatches 15; Indels 8; Gaps 1;

```
Qy           5 GCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL 50
               |   | |||: : | : || :   ||         |:|:|||  |:
Db         203 GNEGQGISPLMAESADQLVHISMKGQ--------AESLNVAVAAGI 240
```

RESULT 38
US-10-072-621-10
; Sequence 10, Application US/10072621

; Publication No. US20020169137A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Connop, Bruce P.
; APPLICANT: Pollard, Michelle
; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
; TITLE OF INVENTION: BY MODIFICATION OF ABC TRANSPORTER EXPRESSION OR
ACTIVITY
; FILE REFERENCE: 100103.402
; CURRENT APPLICATION NUMBER: US/10/072,621
; CURRENT FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-621-10

Query Match 13.9%; Score 59; DB 14; Length 638;
Best Local Similarity 28.2%; Pred. No. 2e+02;
Matches 22; Conservative 9; Mismatches 33; Indels 14; Gaps 3;

```
Qy          14 MRSISENSLVAMDFSGQKSRVIEN-PTEALSVAVEEGLAWRKKG--CLRLGTHG------ 64
               :: :  |   |  ||    |   |     || :| ||  |||||   |  |  |
Db          13 LKKVDNNLTEAQRFSSLPRRAAVNIEFRDLSYSVPEGPWWRKKGYKTLLKGISGKFNSGE 72

Qy          65 -----SPTASSQSSATNM 77
                     |: : :|:  |:
Db          73 LVAIMGPSGAGKSTLMNI 90
```

RESULT 39
US-10-211-962-85
; Sequence 85, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 85
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-10-211-962-85

Query Match 13.9%; Score 59; DB 15; Length 1024;

```
  Best Local Similarity   27.7%;  Pred. No. 3.7e+02;
  Matches   13;  Conservative   14;  Mismatches   18;  Indels    2;  Gaps    1;

Qy         15 RSISENSLVAMDFSGQ--KSRVIENPTEALSVAVEEGLAWRKKGCLR 59
              :::: ||  : |||||     ::|: :||  :: :    || :  :|
Db        460 QAVAANSAASRDFSGQGGLGELLESRSEASKLSSKTAKEWRNRRKVR 506


RESULT 40
US-10-104-047-3196
; Sequence 3196, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3196
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3196

  Query Match             13.8%;  Score 58.5;  DB 12;  Length 189;
  Best Local Similarity   28.8%;  Pred. No. 49;
  Matches   19;  Conservative    9;  Mismatches   21;  Indels   17;  Gaps    2;

Qy         36 ENPTEALSVA----------------VEEGLAWRKKGCLRLGTHGS-PTASSQSSATNMA 78
              ::|  |  ||                || |     :  | | | |  ||::|||:   :
Db         24 DSPASASRVAGTTGTRHHAQLIFVFLVETGFRHIGQAALELLTSGDPPTSASQSAGITVL 83

Qy         79 IHRSQP 84
               ||::|
Db         84 SHRTRP 89


Search completed: January 13, 2004, 16:32:03
Job time : 37.378 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:14:47 ; Search time 42.3307 Seconds
(without alignments)
512.073 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

1	383	90.5	207	11	Q8VDI2	Q8vdi2 mus musculu
2	188	44.4	541	11	Q91WC5	Q91wc5 mus musculu
3	188	44.4	596	11	Q8BSS5	Q8bss5 mus musculu
4	188	44.4	596	11	Q8BSH4	Q8bsh4 mus musculu
5	186	44.0	535	11	Q9QZC5	Q9qzc5 rattus norv
6	168.5	39.8	447	4	Q9Y220	Q9y220 homo sapien
7	76	18.0	1344	3	Q8WZS4	Q8wzs4 neurospora
8	74.5	17.6	655	10	Q9C620	Q9c620 arabidopsis
9	70.5	16.7	346	16	Q8U8L9	Q8u8l9 agrobacteri
10	70	16.5	621	4	Q9BUJ3	Q9buj3 homo sapien
11	70	16.5	1664	4	Q9BZE5	Q9bze5 homo sapien
12	69	16.3	554	10	Q8LQB2	Q8lqb2 oryza sativ
13	68.5	16.2	533	11	Q9Z2Z2	Q9z2z2 mus musculu
14	68.5	16.2	545	4	Q96JP3	Q96jp3 homo sapien
15	68.5	16.2	642	17	Q8PUS8	Q8pus8 methanosarc
16	68.5	16.2	686	11	Q8C208	Q8c208 mus musculu
17	68.5	16.2	868	10	Q9SH67	Q9sh67 arabidopsis
18	68.5	16.2	1664	13	Q8JIF9	Q8jif9 acanthogobi
19	68	16.1	653	16	Q9JSF0	Q9jsf0 chlamydia p
20	68	16.1	1240	3	Q9P6U5	Q9p6u5 neurospora
21	66	15.6	642	17	Q8PYV1	Q8pyv1 methanosarc
22	66	15.6	667	2	Q44062	Q44062 aeromonas h
23	65.5	15.5	455	2	Q8GJN3	Q8gjn3 synechococc
24	65.5	15.5	658	16	Q8DW01	Q8dw01 streptococc
25	65.5	15.5	899	2	Q8KJE6	Q8kje6 rhizobium l
26	65	15.4	612	5	O17206	O17206 caenorhabdi
27	65	15.4	653	16	Q9Z8C4	Q9z8c4 chlamydia p
28	65	15.4	786	12	Q8V3L5	Q8v3l5 swinepox vi
29	65	15.4	1275	4	Q9UQ36	Q9uq36 homo sapien
30	65	15.4	1313	2	Q93UN0	Q93un0 helicobacte
31	65	15.4	1783	4	O15038	O15038 homo sapien
32	65	15.4	1791	4	O60382	O60382 homo sapien
33	65	15.4	2296	4	Q9UHA8	Q9uha8 homo sapien
34	65	15.4	2752	4	Q9UQ35	Q9uq35 homo sapien
35	64.5	15.2	256	10	Q9M210	Q9m210 arabidopsis
36	64.5	15.2	681	11	Q8VIM3	Q8vim3 mus musculu
37	64.5	15.2	689	11	Q91ZE5	Q91ze5 mus musculu
38	64.5	15.2	689	11	Q8BYX0	Q8byx0 mus musculu
39	64.5	15.2	733	4	Q9UBZ1	Q9ubz1 homo sapien
40	64.5	15.2	1004	17	Q8TJS3	Q8tjs3 methanosarc
41	64.5	15.2	1677	5	Q9BKV5	Q9bkv5 leishmania
42	64.5	15.2	2303	4	O95996	O95996 homo sapien
43	64	15.1	313	17	Q9YAQ7	Q9yaq7 aeropyrum p
44	64	15.1	470	16	Q8XB83	Q8xb83 escherichia
45	64	15.1	719	11	Q91YW8	Q91yw8 mus musculu

ALIGNMENTS

RESULT 1
Q8VDI2
ID Q8VDI2 PRELIMINARY; PRT; 207 AA.
AC Q8VDI2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```
DE   Similar to growth factor receptor-bound protein 10 (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021820; AAH21820.1; -.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   Receptor.
FT   NON_TER       1      1
SQ   SEQUENCE   207 AA;  23393 MW;  02D0C5231D884882 CRC64;

  Query Match             90.5%;  Score 383;  DB 11;  Length 207;
  Best Local Similarity   86.9%;  Pred. No. 1.2e-36;
  Matches   73;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              |||| |:|||:||||||:||||||||||||:||||||:||||||||||||||||||||||
Db         22 QGRSACNSQSMSPMRSVSENSLVAMDFSGEKSRVIDNPTEALSVAVEEGLAWRKKGCLRL 81

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              | ||||:| ||||| |||:|||||
Db         82 GNHGSPSAPSQSSAVNMALHRSQP 105


RESULT 2
Q91WC5
ID   Q91WC5      PRELIMINARY;      PRT;   541 AA.
AC   Q91WC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to growth factor receptor bound protein 10.
GN   GRB10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye, and Retina;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR   EMBL; BC016111; AAH16111.1; -.
DR   MGD; MGI:103232; Grb10.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
```

```
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   Receptor.
SQ   SEQUENCE   541 AA;  61217 MW;  A8FA9ED57C85F674 CRC64;

  Query Match             44.4%;  Score 188;  DB 11;  Length 541;
  Best Local Similarity   54.1%;  Pred. No. 2.3e-13;
  Matches   46;  Conservative    7;  Mismatches   22;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||  |||:|| || | |:||| ||||:   |:
Db        360 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWRKRS-TRMN- 417

Qy         63 HGSPTASSQS----SATNMAIHRSQ 83
                  ||||    |  |  |||:|
Db        418 ----ILSSQSPLHPSTLNAVIHRTQ 438


RESULT 3
Q8BSS5
ID   Q8BSS5      PRELIMINARY;      PRT;   596 AA.
AC   Q8BSS5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Growth factor receptor bound protein 10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030727; BAC27100.1; -.
SQ   SEQUENCE   596 AA;  67543 MW;  EB13CA896DF41533 CRC64;

  Query Match             44.4%;  Score 188;  DB 11;  Length 596;
  Best Local Similarity   54.1%;  Pred. No. 2.6e-13;
  Matches   46;  Conservative    7;  Mismatches   22;  Indels   10;  Gaps    3;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              | |      :||||:||||||||||||  |||:|| || | |:||| ||||:   |:
Db        415 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWRKRS-TRMN- 472
```

```
Qy          63 HGSPTASSQS----SATNMAIHRSQ 83
                     ||||    |  |  |||:|
Db         473 ----ILSSQSPLHPSTLNAVIHRTQ 493


RESULT 4
Q8BSH4
ID   Q8BSH4      PRELIMINARY;      PRT;   596 AA.
AC   Q8BSH4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Growth factor receptor bound protein 10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Mesonephros;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK032927; BAC28088.1; -.
SQ   SEQUENCE   596 AA;  67573 MW;  EB13D6E51DE87943 CRC64;

  Query Match             44.4%;  Score 188;  DB 11;  Length 596;
  Best Local Similarity   54.1%;  Pred. No. 2.6e-13;
  Matches   46;  Conservative    7;  Mismatches   22;  Indels   10;  Gaps    3;

Qy           3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
               | |      :||||:|||||||||||||  |||:|| || | |:||| ||||:   |:
Db         415 RKGLPPPFNAPMRSVSENSLVAMDFSGQIGRVIDNPAEAQSAALEEGHAWRKRS-TRMN- 472

Qy          63 HGSPTASSQS----SATNMAIHRSQ 83
                     ||||    |  |  |||:|
Db         473 ----ILSSQSPLHPSTLNAVIHRTQ 493


RESULT 5
Q9QZC5
ID   Q9QZC5      PRELIMINARY;      PRT;   535 AA.
AC   Q9QZC5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Growth factor receptor binding protein GRB7.
GN   GRB7.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=98421528; PubMed=9748281;
RA   Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., Van Obberghen E.,
RA   Mauvais-Jarvis F., Girard J., Burnol A.F.;
RT   "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT   actions.";
RL   J. Biol. Chem. 273:26026-26035(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20260602; PubMed=10803466;
RA   Kasus-Jacobi A., Bereziat V., Perdereau D., Girard J., Burnol A.F.;
RT   "Evidence for an interaction between the insulin receptor and Grb7. A
RT   role for two of its binding domains, PIR and SH2.";
RL   Oncogene 19:2052-2059(2000).
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR   EMBL; AF190121; AAF01776.1; -.
DR   HSSP; P35235; 1AYA.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   Receptor.
SQ   SEQUENCE   535 AA;  59889 MW;  15DB67C4D19B89E4 CRC64;

  Query Match           44.0%;  Score 186;  DB 11;  Length 535;
  Best Local Similarity  59.7%;  Pred. No. 3.9e-13;
  Matches   43;  Conservative    6;  Mismatches   19;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   ||||||  ||||  | ||  |||||   ||      ||  |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIENPQEALSAATEEAQAWRKKTNHRLSL---PTPCSGL 422

Qy         73 SATNMAIHRSQP 84
              |  : ||||:||
Db        423 S-LSAAIHRTQP 433


RESULT 6
Q9Y220
ID   Q9Y220      PRELIMINARY;      PRT;   447 AA.
AC   Q9Y220;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Grb7V protein.
GN   GRB7V.
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98376491; PubMed=9710451;
RA   Tanaka S., Mori M., Akiyoshi T., Tanaka Y., Mafune K., Wands J.R.,
RA   Sugimachi K.;
RT   "A novel variant of human Grb7 associated with invasive esophageal
RT   carcinoma.";
RL   J. Clin. Invest. 102:821-827(1998).
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR   EMBL; AB008790; BAA29060.1; -.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
SQ   SEQUENCE   447 AA;  49506 MW;  EC87F21A1C6439D5 CRC64;

  Query Match          39.8%;  Score 168.5;  DB 4;  Length 447;
  Best Local Similarity   51.2%;  Pred. No. 3.5e-11;
  Matches   42;  Conservative    5;  Mismatches   16;  Indels   19;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL-------GT--- 62
              |:|| |:|:||||||||   |||||| |||||||:||  |||||   ||        ||
Db        363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKKTNHRLSLPMPASGTSLS 422

Qy         63 ---------HGSPTASSQSSAT 75
                      |:| | |   ||
Db        423 AACSWSGRVSGTPRALSSLCAT 444


RESULT 7
Q8WZS4
ID   Q8WZS4      PRELIMINARY;      PRT;  1344 AA.
AC   Q8WZS4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 138.9 kDa protein.
GN   B8L21.130.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
```

```
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL669989; CAD21099.1; -.
DR   InterPro; IPR000910; HMG_12_box.
DR   Pfam; PF00505; HMG_box; 1.
DR   SMART; SM00398; HMG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1344 AA;  138944 MW;  B1AB8BF7527081EE CRC64;

  Query Match          18.0%;  Score 76;  DB 3;  Length 1344;
  Best Local Similarity   30.1%;  Pred. No. 8.5;
  Matches   25;  Conservative   12;  Mismatches   22;  Indels   24;  Gaps    3;

Qy          4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTH 63
              |  || | ||:      :|:: | | || |          :::: |||
Db        172 SSSSSNSSSPLTRKRAATLISTDLSSQKPR----------LSIDPGLA----------G 210

Qy         64 GSPTASSQSSATNMA---IHRSQ 83
              |: | :||| :|  |   || :|
Db        211 GAATGASQSRSTTTAAESIHNAQ 233


RESULT 8
Q9C620
ID   Q9C620      PRELIMINARY;      PRT;   655 AA.
AC   Q9C620;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Receptor serine/threonine kinase PR5K, putative.
GN   T4O24.8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
```

```
RL   Nature 408:816-820(2000).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AC083891; AAG50590.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   655 AA;  73013 MW;  7808804B621A9566 CRC64;

  Query Match             17.6%;  Score 74.5;  DB 10;  Length 655;
  Best Local Similarity   25.6%;  Pred. No. 5.3;
  Matches   23;  Conservative   16;  Mismatches   34;  Indels   17;  Gaps    3;

Qy          11 ISPMRSISENSLVAMDFSGQKSRVIENP--------------TEALSVAVEEGLAWRKKG 56
               : |   :  || :  || |  || : ||               |: | :|:|:| |
Db         164 LPPSLKLEGNSFLLNDFGGSCSRNVSNPASRTALNTLESTPSTDNLKIALEDGFALEVNS 223

Qy          57 CLR--LGTHGSPTASSQSSATNMAIHRSQP 84
                 |  : : |:    ||:|:  :  :| :|
Db         224 DCRTCIDSKGA-CGFSQTSSRFVCYYRQEP 252


RESULT 9
Q8U8L9
ID   Q8U8L9      PRELIMINARY;      PRT;   346 AA.
AC   Q8U8L9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu4071.
GN   ATU4071 OR AGR_L_1570.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009338; AAL44872.1; -.
DR   EMBL; AE008277; AAK89358.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   346 AA;  37882 MW;  6EC2B813564FD385 CRC64;

  Query Match             16.7%;  Score 70.5;  DB 16;  Length 346;
  Best Local Similarity   27.9%;  Pred. No. 7.1;
  Matches   24;  Conservative   13;  Mismatches   30;  Indels   19;  Gaps    4;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA--------WRK 54
              |:||    : |  |     |::  :: |  | :|    || :||| ||         | |
Db        194 RAGCDLNPLDPSSSEDRLRLMSYIWADQTDR-LERTAAALRIAVENGLQVEKADAVDWLK 252

Qy         55 KGCLRLGTHGSPTASSQSSATNMAIH 80
              :   ||       |:  : ||::  |
Db        253 R---RL-------ATQHTGATHVVYH 268


RESULT 10
Q9BUJ3
ID   Q9BUJ3      PRELIMINARY;      PRT;   621 AA.
AC   Q9BUJ3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002561; AAH02561.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   621 AA;  67813 MW;  3DA0D4A18D3A2466 CRC64;

  Query Match             16.5%;  Score 70;  DB 4;  Length 621;
  Best Local Similarity   25.8%;  Pred. No. 17;
  Matches   24;  Conservative   17;  Mismatches   32;  Indels   20;  Gaps    3;
```

```
Qy          1 QGRSGCSSQSISP----MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG 56
              ||| | :|:|:|         | :|  :   |  :|| :  | :           ||:
Db        387 QGRRGRNSRSVSSGSNRTSEASSSSSSSSSSRSRSRSLSPPHK----------RWRRSS 436

Qy         57 C------LRLGTHGSPTASSQSSATNMAIHRSQ 83
              |       |  :  | ::|| ||::: :  ||:
Db        437 CSSSGRSRRCSSSSSSSSSSSSSSSSSSRSR 469


RESULT 11
Q9BZE5
ID   Q9BZE5      PRELIMINARY;      PRT;  1664 AA.
AC   Q9BZE5; Q9Y4E0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   PGC-1 related co-activator.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21238514; PubMed=11340167;
RA   Andersson U., Scarpulla R.C.;
RT   "Pgc-1-related coactivator, a novel, serum-inducible coactivator of
RT   nuclear respiratory factor 1-dependent transcription in mammalian
RT   cells.";
RL   Mol. Cell. Biol. 21:3738-3749(2001).
DR   EMBL; AF325193; AAK11573.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1664 AA;  177666 MW;  8AF8E83D2A1C89FB CRC64;

  Query Match             16.5%;  Score 70;  DB 4;  Length 1664;
  Best Local Similarity   25.8%;  Pred. No. 55;
  Matches   24;  Conservative   17;  Mismatches   32;  Indels   20;  Gaps    3;

Qy          1 QGRSGCSSQSISP----MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG 56
              ||| | :|:|:|         | :|  :   |  :|| :  | :           ||:
Db       1420 QGRRGRNSRSVSSGSNRTSEASSSSSSSSSSRSRSRSLSPPHK----------RWRRSS 1469

Qy         57 C------LRLGTHGSPTASSQSSATNMAIHRSQ 83
              |       |  :  | ::|| ||::: :  ||:
Db       1470 CSSSGRSRRCSSSSSSSSSSSSSSSSSSRSR 1502


RESULT 12
Q8LQB2
ID   Q8LQB2      PRELIMINARY;      PRT;   554 AA.
AC   Q8LQB2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
```

```
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative potassium-sodium symporter.
GN   OSJNBB0022N24.16.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OSJNBb0022N24.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003567; BAB93392.1; -.
DR   Gramene; Q8LQB2; -.
DR   InterPro; IPR003445; Cat_transpt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF02386; TrkH; 2.
DR   PROSITE; PS00334; MYB_2; 1.
SQ   SEQUENCE   554 AA;  60218 MW;  5433B2BB030F2ACB CRC64;

  Query Match           16.3%;  Score 69;  DB 10;  Length 554;
  Best Local Similarity  32.4%;  Pred. No. 19;
  Matches   24;  Conservative   13;  Mismatches   33;  Indels    4;  Gaps    3;

Qy          5 GCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSV--AVEEGLAWR-KKGCLRLG 61
              | | :   |  | | ::|| :: :     |: |||   :    || ||  | |:||
Db        132 GGSGKPPPPTTSPS-STLVELELAPPMDVVVVNPTTTATTHDEVELGLGRRNKRGCTCTT 190

Qy         62 THGSPTASSQSSAT 75
              || | ::|:  : |
Db        191 THTSSSSSASKTTT 204


RESULT 13
Q9Z2Z2
ID   Q9Z2Z2      PRELIMINARY;      PRT;   533 AA.
AC   Q9Z2Z2;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Eos protein.
GN   ZNFN1A4 OR EOS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE=99232954; PubMed=10218586;
RA   Homma Y., Kiyosawa H., Mori T., Oguri A., Nikaido T., Kanazawa K.,
RA   Tojo M., Takeda J., Tanno Y., Yokoya S., Kawabata I., Ikeda H.,
RA   Wanaka A.;
```

```
RT   "Eos: a novel member of the Ikaros gene family expressed predominantly
RT   in the developing nervous system.";
RL   FEBS Lett. 447:76-80(1999).
DR   EMBL; AB017615; BAA36213.1; -.
DR   HSSP; P15822; 1BBO.
DR   MGD; MGI:1343139; Znfn1a4.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   533 AA;  58167 MW;  7A5FF32C6FFDC372 CRC64;

  Query Match             16.2%;  Score 68.5;  DB 11;  Length 533;
  Best Local Similarity   38.0%;  Pred. No. 21;
  Matches   19;  Conservative    9;  Mismatches   17;  Indels    5;  Gaps    1;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 51
              :||  || ||:| ||:    :| ::|  :  | | |     || ||: |:
Db         31 NSQHSSPSRSLSANSIKVEMYSDEESSRLLGPDERLLDKDDSVIVEDSLS 80


RESULT 14
Q96JP3
ID   Q96JP3      PRELIMINARY;      PRT;   545 AA.
AC   Q96JP3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein KIAA1782 (Fragment).
GN   KIAA1782.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large Proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058685; BAB47411.1; -.
DR   Genew; HGNC:13179; ZNFN1A4.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   545 AA;  59742 MW;  7A8539E5B8F9BD84 CRC64;
```

```
  Query Match             16.2%;  Score 68.5;  DB 4;  Length 545;
  Best Local Similarity   38.0%;  Pred. No. 21;
  Matches   19;  Conservative    9;  Mismatches   17;  Indels    5;  Gaps    1;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 51
              :||  || ||:| ||:    :| ::|  :  | | |     || ||: |:
Db         44 NSQHSSPSRSLSANSIKVEMYSDEESSRLLGPDERLLEKDDSVIVEDSLS 93


RESULT 15
Q8PUS8
ID   Q8PUS8      PRELIMINARY;      PRT;   642 AA.
AC   Q8PUS8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Dihydropyrimidinase (EC 3.5.2.2).
GN   MM2253.
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX   MEDLINE=22120827; PubMed=12125824;
RA   Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA   Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
RA   Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA   Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA   Fritz H.-J., Gottschalk G.;
RT   "The genome of Methanosarcina mazei: evidence for lateral gene
RT   transfer between Bacteria and Archaea.";
RL   J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR   EMBL; AE013466; AAM31949.1; -.
DR   InterPro; IPR002821; Hydantoinase_A.
DR   Pfam; PF01968; Hydantoinase_A; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   642 AA;  70251 MW;  C0C6C23A3B6493B4 CRC64;

  Query Match             16.2%;  Score 68.5;  DB 17;  Length 642;
  Best Local Similarity   31.8%;  Pred. No. 26;
  Matches   28;  Conservative   13;  Mismatches   26;  Indels   21;  Gaps    6;

Qy         13 PMRSISENSLVAMDFSGQ------KSRVIE----NPTEALSVAVEEGLAWRKK----GCL 58
              |:     ::|   ||  |       | |:::     ||:|| | : |  |||::    |
Db        385 PVSVFEISALTRKDFHPQTLDCLIKKRLVQVIGFTPTDALHV-LGEYTAWREEASRTGAE 443

Qy         59 RLG--THGSP----TASSQSSATNMAIH 80
              |||     :|    ||  :  | |||:|
Db        444 RLGRLMRMTPIEFCTAVKKKVARNMALH 471


RESULT 16
Q8C208
ID   Q8C208      PRELIMINARY;      PRT;   686 AA.
```

```
AC   Q8C208;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zinc finger protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK089522; BAC40912.1; -.
SQ   SEQUENCE   686 AA;  75078 MW;  F99ADB635184FAC0 CRC64;

  Query Match             16.2%;  Score 68.5;  DB 11;  Length 686;
  Best Local Similarity   38.0%;  Pred. No. 28;
  Matches   19;  Conservative    9;  Mismatches   17;  Indels    5;  Gaps    1;

Qy          7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 51
              :||  || ||:| ||:    :| ::|  :  | | |     || ||: |:
Db         84 NSQHSSPSRSLSANSIKVEMYSDEESSRLLGPDERLLDKDDSVIVEDSLS 133


RESULT 17
Q9SH67
ID   Q9SH67      PRELIMINARY;      PRT;   868 AA.
AC   Q9SH67;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   F22C12.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Khan S., Brooks S., Buehler E., Chao Q., Dunn P., Kim C.,
RA   Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA   Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA   Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA   Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome
RT   I.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007764; AAF24561.1; -.
DR   InterPro; IPR006153; Na_H_porter.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
SQ   SEQUENCE   868 AA;  94617 MW;  4394523B169E6979 CRC64;
```

```
  Query Match             16.2%;  Score 68.5;  DB 10;  Length 868;
  Best Local Similarity   30.7%;  Pred. No. 37;
  Matches   23;  Conservative   11;  Mismatches   26;  Indels   15;  Gaps    3;

Qy          4 SGCSSQSISPM----RSI-SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCL 58
              | |::  | |     ||:  |   |  :: |   ||:||   :: : |: |
Db        639 SKCTAFVILPFHKQWRSLEKEFETVRSEYQGINKRVLENSPCSVGILVDRG--------- 689

Qy         59 RLGTHGSPTASSQSS 73
               || : || |||  |
Db        690 -LGDNNSPVASSNFS 703


RESULT 18
Q8JIF9
ID   Q8JIF9      PRELIMINARY;      PRT;  1664 AA.
AC   Q8JIF9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Vitellogenin.
GN   VG-530.
OS   Acanthogobius flavimanus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidei;
OC   Gobiidae; Acanthogobius.
OX   NCBI_TaxID=86203;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkubo N., Mochida K., Adachi S., Hara A., Matsubara T.;
RT   "Deduced primary structures of two form of vitellogenin in Japanese
RT   common goby.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088473; BAC06190.1; -.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF01347; Vitellogenin_N; 1.
DR   Pfam; PF00094; vwd; 1.
DR   SMART; SM00638; LPD_N; 1.
DR   SMART; SM00216; VWD; 1.
SQ   SEQUENCE   1664 AA;  185650 MW;  1A2909403485578A CRC64;

  Query Match             16.2%;  Score 68.5;  DB 13;  Length 1664;
  Best Local Similarity   29.8%;  Pred. No. 83;
  Matches   25;  Conservative   14;  Mismatches   30;  Indels   15;  Gaps    3;

Qy          1 QGRSGCSSQSISPMRSISENSLV-AMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLR 59
              | |:  || | |  ||:  ||   :   |  :|:|     :|:             | :
Db       1063 QNRTSSSSSS-SSSRSVLRNSRTSSSSSSSSRSKVTSKVIKAM-----------GKIL 1108

Qy         60 LGTHGSPTASSQSSATNMAIHRSQ 83
               |:| | ::|| ||:::  | | |
Db       1109 GGSHKSSSSSSSSSSSRRISRQQ 1132
```

```
RESULT 19
Q9JSF0
ID   Q9JSF0      PRELIMINARY;      PRT;   653 AA.
AC   Q9JSF0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transglycolase/transpeptidase.
GN   PBP3.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
DR   EMBL; AP002546; BAA98627.1; -.
DR   InterPro; IPR005311; PBP_dimer.
DR   InterPro; IPR001460; Transpeptdse.
DR   Pfam; PF03717; PBP_dimer; 1.
DR   Pfam; PF00905; Transpeptidase; 1.
SQ   SEQUENCE   653 AA;  73619 MW;  3CD2334EEFA0979C CRC64;

  Query Match             16.1%;  Score 68;  DB 16;  Length 653;
  Best Local Similarity   30.5%;  Pred. No. 30;
  Matches   25;  Conservative   14;  Mismatches   39;  Indels    4;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGT 62
              |: |  :  ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db        353 RTLCPGRKGSPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG- 411

Qy         63 HGSPTA---SSQSSATNMAIHR 81
                |  |     |::|    : ||
Db        412 FGRKTGIELPSEASGLVPSPHR 433


RESULT 20
Q9P6U5
ID   Q9P6U5      PRELIMINARY;      PRT;  1240 AA.
AC   Q9P6U5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Related to protease ULP2 protein.
GN   15E6.80.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
```

```
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353822; CAB88639.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Protease.
SQ   SEQUENCE   1240 AA;  138114 MW;  716E38F4DF0D177A CRC64;

  Query Match             16.1%;  Score 68;  DB 3;  Length 1240;
  Best Local Similarity   34.4%;  Pred. No. 66;
  Matches   22;  Conservative    5;  Mismatches   23;  Indels   14;  Gaps    2;

Qy         32 SRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTA----------SSQSSATNMA 78
              |||    | || |     :|:   | :   | |||||             ||| : :|
Db        386 SRVTRT-TSALDVEGSRNMAFEPAGLIAQATAGSPTASTRRRPRLVDTLLSSQQALSNQY 444

Qy         79 IHRS 82
              |||
Db        445 EHRS 448


RESULT 21
Q8PYV1
ID   Q8PYV1      PRELIMINARY;      PRT;   642 AA.
AC   Q8PYV1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Dihydropyrimidinase (EC 3.5.2.2).
GN   MM0750.
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX   MEDLINE=22120827; PubMed=12125824;
RA   Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA   Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
RA   Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA   Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA   Fritz H.-J., Gottschalk G.;
RT   "The genome of Methanosarcina mazei: evidence for lateral gene
RT   transfer between Bacteria and Archaea.";
RL   J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR   EMBL; AE013300; AAM30446.1; -.
DR   InterPro; IPR002821; Hydantoinase_A.
DR   Pfam; PF01968; Hydantoinase_A; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   642 AA;  69827 MW;  758FFE70478103A8 CRC64;
```

```
  Query Match             15.6%;  Score 66;  DB 17;  Length 642;
  Best Local Similarity   26.1%;  Pred. No. 51;
  Matches   29;  Conservative   18;  Mismatches   30;  Indels   34;  Gaps    6;

Qy          3 RSGCSSQSISPMRS----------ISEN---SLVAMDFSGQ------KSRVIE----NPT 39
              ||| ::  || : |          :| |   :|:  |   |      | |:|:     ||
Db        362 RSGYTAGEISKVESEVLGVIGDEPVSVNDIKTLIRKDLHPQTLDSLIKKRLIQAIGFTPT 421

Qy         40 EALSVAVEEGLAWRKKG----------CLRLGTHGSPTASSQSSATNMAIH 80
              :|| | : |  || ::           :|:  |   |:  :  | ||::|
Db        422 DALHV-LGEYTAWNEEASRIGAERLARLMRMTPHEFCTSVKKKVARNMSLH 471


RESULT 22
Q44062
ID   Q44062      PRELIMINARY;      PRT;   667 AA.
AC   Q44062;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Amylase.
GN   AMYB.
OS   Aeromonas hydrophila.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
OC   Aeromonadaceae; Aeromonas.
OX   NCBI_TaxID=644;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JMP636;
RA   Kidd S.P., Pemberton J.M.;
RT   "Aeromonas hydrophila amyB.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77866; AAA98043.1; -.
DR   HSSP; P29957; 1AQM.
DR   InterPro; IPR006048; Alpha_amyl_C.
DR   InterPro; IPR006047; Alpha_amyl_cat.
DR   Pfam; PF00128; alpha-amylase; 1.
DR   Pfam; PF02806; alpha-amylase_C; 1.
SQ   SEQUENCE   667 AA;  72719 MW;  2CEFB8B086774DA6 CRC64;

  Query Match             15.6%;  Score 66;  DB 2;  Length 667;
  Best Local Similarity   29.9%;  Pred. No. 53;
  Matches   20;  Conservative    8;  Mismatches   27;  Indels   12;  Gaps    3;

Qy          2 GRSGCSSQSISPMRSISENSLV-----AMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG 56
              | || |  |: | | :     :      |  |:||  | ::   | | |: |
Db        276 GESGASGHSLQPFRPVHRLGTIGTVFTAASFNGQ-FRNLKTKAERLGVSAE------IHA 328

Qy         57 CLRLGTH 63
              |  ||:|
Db        329 CTNLGSH 335


RESULT 23
Q8GJN3
ID   Q8GJN3      PRELIMINARY;      PRT;   455 AA.
AC   Q8GJN3;
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MurF (EC 6.3.2.15).
GN   SEM0006.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Holtman C.K., Sandoval P., Chen Y., Socias T., McMurtry S.,
RA   Gonzalez A., Salinas I., Golden S.S., Youderian P.;
RT   "Synechococcus elongatus PCC7942 cosmid 4G8.";
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY157498; AAN46171.1; -.
KW   Ligase.
SQ   SEQUENCE   455 AA;  48966 MW;  F7ABCF0E46AD3D8E CRC64;

  Query Match           15.5%;  Score 65.5;  DB 2;  Length 455;
  Best Local Similarity   31.6%;  Pred. No. 38;
  Matches   24;  Conservative   10;  Mismatches   21;  Indels   21;  Gaps    5;

Qy         19 ENSLVAMD-FSGQKSRVIENPTEALSVAVEEGL----------AWRKKGCL---RLGTHG 64
              |: | |:  | |        ||   |   :||:          :||| |||   ||  :
Db        337 ESMLAALQAFGG-----YAGPTPDCSAGHDEGIGRFQRNLPSPSWRKSGCLGLDRLLIYA 391

Qy         65 SPT--ASSQSSATNMA 78
               ||  |: |: |: :|
Db        392 DPTEAAAMQAGASAIA 407


RESULT 24
Q8DW01
ID   Q8DW01      PRELIMINARY;      PRT;   658 AA.
AC   Q8DW01;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Transketolase (EC 2.2.1.1).
GN   TKT OR SMU.291.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UA159 / ATCC 700610 / Serotype C;
RX   MEDLINE=22295063; PubMed=12397186;
RA   Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA   Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA   Li S., Zhu H., Najar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT   "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT   pathogen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR   EMBL; AE014878; AAN58055.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   658 AA;  71075 MW;  0A996A8DAFCAB68C CRC64;
```

```
  Query Match             15.5%;  Score 65.5;  DB 16;  Length 658;
  Best Local Similarity   23.2%;  Pred. No. 60;
  Matches   19;  Conservative   18;  Mismatches   28;  Indels   17;  Gaps    2;

Qy         16 SISENSLVAMDFSGQKSRVIE--NPTEALSVAVEEGLAWRKKGCLRLGT---------- 62
              : :|:     |  |  : ::|  |  ||: :|:||  |  |   : : |
Db        195 AFTESVRARYDAYGWHTILVEDGNNIEAIGLAIEEAKAAGKPSLIEIKTVIGYGAPTKGG 254

Qy         63 ----HGSPTASSQSSATNMAIH 80
                  ||:|  : :::||  |::
Db        255 TNAVHGAPLGAEEAAATRKALN 276


RESULT 25
Q8KJE6
ID   Q8KJE6      PRELIMINARY;      PRT;   899 AA.
AC   Q8KJE6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Fusion protein CONTAINS putative ligase and probable ARGINOSUCCINATE
DE   lyase.
GN   MSI203.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R7A;
RX   MEDLINE=21999272; PubMed=12003951;
RA   Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,
RA   Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,
RA   Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;
RT   "Comparative sequence analysis of the symbiosis island of
RT   Mesorhizobium loti strain R7A.";
RL   J. Bacteriol. 184:3086-3095(2002).
DR   EMBL; AL672113; CAD31608.1; -.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR000362; Fumarate_lyase.
DR   Pfam; PF00206; lyase_1; 1.
DR   PRINTS; PR00149; FUMRATELYASE.
DR   PROSITE; PS00867; CPSASE_2; 1.
SQ   SEQUENCE   899 AA;  97088 MW;  092265C652341D81 CRC64;

  Query Match             15.5%;  Score 65.5;  DB 2;  Length 899;
  Best Local Similarity   28.0%;  Pred. No. 88;
  Matches   23;  Conservative   12;  Mismatches   36;  Indels   11;  Gaps    2;

Qy          5 GCSSQSISP---MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
              |||  |::     |:|   ||:    |   | ::||        :|:|||       |:
Db        739 GCSPISLAEGALKRAIILTSLIVKFMSFNVSAMLEN--------LEDGLAMTTVAAERMA 790

Qy         62 THGSPTASSQSSATNMAIHRSQ 83
                | |  |: :    :|   ||
Db        791 VRGVPFRSAHTQIGEIAARLSQ 812
```

```
RESULT 26
O17206
ID   O17206      PRELIMINARY;      PRT;   612 AA.
AC   O17206;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   C01B12.3 protein.
GN   C01B12.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Scheet P., Maggi L.;
RT   "The sequence of C. elegans cosmid C01B12.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF025458; AAB70976.1; -.
DR   WormPep; C01B12.3; CE07791.
DR   InterPro; IPR000615; Worm_fam_8.
DR   Pfam; PF01062; DUF289; 1.
DR   ProDom; PD002802; Worm_fam_8; 1.
SQ   SEQUENCE   612 AA;  71031 MW;  DFBB43916541DD44 CRC64;

  Query Match             15.4%;  Score 65;  DB 5;  Length 612;
  Best Local Similarity   28.7%;  Pred. No. 63;
  Matches   29;  Conservative    8;  Mismatches   30;  Indels   34;  Gaps    4;

Qy         10 SISPMRSISE-------NSLVAMDFSGQKSRVIENPT--------------------EAL 42
                 |  |   : |       || |  :  | | | ::|||                     |:|
Db        496 SSMPQTQLEEMLKNKNFNSPVKYNTDGMKDRELQNPTPITDHIDLPLHVASSQSWFNESL 555
```

```
Qy          43 SVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQ 83
                  |  ||  | ||          || :|  ||   |:| ||:
Db         556 PVIKEEEEAKRKSNT----DTESPKSSKHSS---MSIRRSE 589


RESULT 27
Q9Z8C4
ID   Q9Z8C4      PRELIMINARY;      PRT;   653 AA.
AC   Q9Z8C4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TRANSGLYCOLASE/TRANSPEPTIDASE (Penicillin-binding protein).
GN   PBP3 OR CPN0419 OR CP0335.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE001625; AAD18563.1; -.
DR   EMBL; AE002196; AAF38189.1; -.
DR   TIGR; CP0335; -.
DR   InterPro; IPR005311; PBP_dimer.
DR   InterPro; IPR001460; Transpeptdse.
DR   Pfam; PF03717; PBP_dimer; 1.
DR   Pfam; PF00905; Transpeptidase; 1.
KW   Complete proteome.
SQ   SEQUENCE   653 AA;  73663 MW;  F466221FABA75E7B CRC64;

  Query Match             15.4%;  Score 65;  DB 16;  Length 653;
  Best Local Similarity   31.3%;  Pred. No. 68;
  Matches   26;  Conservative   12;  Mismatches   35;  Indels   10;  Gaps    3;

Qy           2 GRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLG 61
               || |      ||:: || || : |  : |||  :     |  :    |:|| ::  | ||
Db         358 GRKG------SPLKDISRNSQLNMYMAIQKSSNVYVAQLADRIIQSLGVAWYQQKLLALG 411

Qy          62 THGSPTA---SSQSSATNMAIHR 81
```

```
                |  |     |::|    : ||
Db        412 -FGRKTGIELPSEASGLVPSPHR 433


RESULT 28
Q8V3L5
ID   Q8V3L5      PRELIMINARY;      PRT;   786 AA.
AC   Q8V3L5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   SPV080 putative NTPase.
GN   SPV080.
OS   Swinepox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Suipoxvirus.
OX   NCBI_TaxID=10276;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=17077-99;
RX   MEDLINE=21624277; PubMed=11752168;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Osorio F.A., Balinsky C.,
RA   Kutish G.F., Rock D.L.;
RT   "The genome of swinepox virus.";
RL   J. Virol. 76:783-790(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=17077-99;
RA   Afonso C.L., Tulman E.R., Lu Z., Balinsky C., Osorio F.A., Zsak L.,
RA   Kutish G.F., Rock D.L.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF410153; AAL69819.1; -.
DR   InterPro; IPR004968; Pox_D5.
DR   Pfam; PF03288; Pox_D5; 1.
SQ   SEQUENCE   786 AA;  90794 MW;  707CDC35D515A985 CRC64;

  Query Match             15.4%;  Score 65;  DB 12;  Length 786;
  Best Local Similarity   32.9%;  Pred. No. 85;
  Matches   28;  Conservative   10;  Mismatches   23;  Indels   24;  Gaps    6;

Qy          7 SSQSISPMRSISENSLVAM-----DFSGQKSRVIENP-TEALSVAVEEGLAWRKKGCLRL 60
              |   | |    : |:|:||     :|    ||  :||   ::  ||     :|:|  ||:
Db        132 SFHMIFPDTYTTMNTLIAMKKPLLEF----SRASDNPLIRSIDTAV-----YRRKATLRI 182

Qy         61 -GTHGSPTASSQSSATNMAIHRSQP 84
               ||  ||        ||  ||  ||
Db        183 VGTRKSP--------TNDKIHIKQP 199


RESULT 29
Q9UQ36
ID   Q9UQ36      PRELIMINARY;      PRT;  1275 AA.
AC   Q9UQ36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   RNA binding protein (Fragment).
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016091; BAA83717.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1275 AA;  136869 MW;  45C2B2F85E98A6F6 CRC64;

  Query Match             15.4%;  Score 65;  DB 4;  Length 1275;
  Best Local Similarity   28.1%;  Pred. No. 1.5e+02;
  Matches   27;  Conservative   12;  Mismatches   35;  Indels   22;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKS----------------RVIENPTEALSVAV 46
              ||  || | |   | | :|  :   ||  |                : : :|| |   ||
Db       1054 RSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 1113

Qy         47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
               ||       | ||:      :  | ::|| ||:::
Db       1114 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 1149


RESULT 30
Q93UN0
ID   Q93UN0      PRELIMINARY;      PRT;  1313 AA.
AC   Q93UN0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   VacA.
GN   VACA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AFN1156;
RA   Ji X.H., Rappuoli R., Telford J.L.;
RT   "Functional analysis of chimeric mutants of the helicobacter pylori
RT   vacA gene.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF191641; AAK56856.1; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR003842; VacA.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF02691; VacA; 1.
DR   PRINTS; PR01656; VACCYTOTOXIN.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
SQ   SEQUENCE   1313 AA;  142077 MW;  F649E2A7E35A6511 CRC64;

  Query Match             15.4%;  Score 65;  DB 2;  Length 1313;
```

```
  Best Local Similarity   27.3%;  Pred. No. 1.6e+02;
  Matches   21;  Conservative   13;  Mismatches   29;  Indels   14;  Gaps    3;

Qy          11 ISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG----------CLRL 60
               |  :|  : | : |   :     :: ||    |    || ||:  |           : |
Db         795 ICVVRKDNLNDIKACGMAIGNQSMVNNPE---SYKYLEGKAWKNTGINKTANNTTIAVNL 851

Qy          61 GTHGSPTASSQSSATNM 77
               | : :|| ||:|: ||:
Db         852 GNNSTPT-SSESNTTNL 867


RESULT 31
O15038
ID   O15038      PRELIMINARY;      PRT;  1783 AA.
AC   O15038;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA0324 (Fragment).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002322; BAA20782.2; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   1783 AA;  190940 MW;  660302F6FD4179AB CRC64;

  Query Match             15.4%;  Score 65;  DB 4;  Length 1783;
  Best Local Similarity   28.1%;  Pred. No. 2.3e+02;
  Matches   27;  Conservative   12;  Mismatches   35;  Indels   22;  Gaps    2;

Qy           3 RSGCSSQSISPMRSISENSLVAMDFSGQKS---------------RVIENPTEALSVAV 46
               ||  || | |   | | :|  :   ||  |                 : : :|| |   ||
Db        1562 RSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 1621

Qy          47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
                ||       | ||:      :  | ::|| ||:::
Db        1622 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 1657


RESULT 32
O60382
ID   O60382      PRELIMINARY;      PRT;  1791 AA.
AC   O60382;
```

```
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA0324 (Fragment).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA   Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA   Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA   Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT   "Sequencing of Human Chromosome 16p13.3.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004493; AAC08453.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   1791 AA;  191306 MW;  3A7B5530AEE95F3E CRC64;

  Query Match             15.4%;  Score 65;  DB 4;  Length 1791;
  Best Local Similarity   28.1%;  Pred. No. 2.3e+02;
  Matches   27;  Conservative   12;  Mismatches   35;  Indels   22;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKS---------------RVIENPTEALSVAV 46
              ||  || | |   | | :|  :   ||  |                : : :|| |   ||
Db       1563 RSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 1622

Qy         47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
               ||       | ||:      :  | ::|| ||:::
Db       1623 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 1658


RESULT 33
Q9UHA8
ID   Q9UHA8      PRELIMINARY;      PRT;  2296 AA.
AC   Q9UHA8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Splicing coactivator subunit SRm300.
GN   SRM300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20132238; PubMed=10668804;
```

```
RA   Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA   Rosonina E., Sharp P.A.;
RT   "The SRm160/300 splicing coactivator subunits.";
RL   RNA 6:111-120(2000).
DR   EMBL; AF201422; AAF21439.1; -.
SQ   SEQUENCE   2296 AA;  251964 MW;  17C0BD4EA10A9CF9 CRC64;

  Query Match           15.4%;  Score 65;  DB 4;  Length 2296;
  Best Local Similarity  31.6%;  Pred. No. 3.1e+02;
  Matches   25;  Conservative   11;  Mismatches   35;  Indels    8;  Gaps    2;

Qy          4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL------AWRKKGC 57
              |  || |     | || |    : || :   : :|| |   || ||       | ||:
Db       2131 SSSSSSSSGSSSSDSEGSSFLCNLSGTEE--VPSPTPAPKEAVREGRPPEPTPAKRKRRS 2188

Qy         58 LRLGTHGSPTASSQSSATN 76
                :  | ::|| ||:::
Db       2189 SSSSSSSSSSSSSSSSSS 2207


RESULT 34
Q9UQ35
ID   Q9UQ35      PRELIMINARY;      PRT;  2752 AA.
AC   Q9UQ35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   RNA binding protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016092; BAA83718.1; -.
DR   Genew; HGNC:16639; SRRM2.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2752 AA;  299672 MW;  109C64F181097123 CRC64;

  Query Match           15.4%;  Score 65;  DB 4;  Length 2752;
  Best Local Similarity  28.1%;  Pred. No. 3.9e+02;
  Matches   27;  Conservative   12;  Mismatches   35;  Indels   22;  Gaps    2;

Qy          3 RSGCSSQSISPMRSISENSLVAMDFSGQKS----------------RVIENPTEALSVAV 46
              ||  || | |   | | :|  :   ||  |                 : : :|| |   ||
Db       2531 RSSSSSSSSSSSSSSSSSSSSSSSSGSSSSDSEGSSLPVQPEVALKRVPSPTPAPKEAV 2590

Qy         47 EEGL------AWRKKGCLRLGTHGSPTASSQSSATN 76
               ||       | ||:      :  | ::|| ||:::
Db       2591 REGRPPEPTPAKRKRRSSSSSSSSSSSSSSSSSSS 2626


RESULT 35
```

```
Q9M210
ID   Q9M210      PRELIMINARY;      PRT;   256 AA.
AC   Q9M210;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Transcription factor-like protein.
GN   T8B10_150.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL138646; CAB81835.1; -.
DR   HSSP; O80337; 2GCC.
DR   InterPro; IPR001471; TF_ERF.
DR   Pfam; PF00847; AP2-domain; 1.
DR   PRINTS; PR00367; ETHRSPELEMNT.
DR   ProDom; PD001423; TF_ERF; 1.
DR   SMART; SM00380; AP2; 1.
SQ   SEQUENCE   256 AA;  28216 MW;  BD9B5CDF3A892A45 CRC64;

  Query Match             15.2%;  Score 64.5;  DB 10;  Length 256;
  Best Local Similarity   32.6%;  Pred. No. 25;
  Matches   29;  Conservative   10;  Mismatches   31;  Indels   19;  Gaps    4;

Qy          7 SSQSI----SPMRSISENSLVAMDFSGQKSRVI------ENPTEALSVAVEEGLAW---- 52
              || |:    |   | |: |||  : || | |        :|||    | :     |
Db         27 SSSSVVTSSSDSWSTSKRSLVQDNDSGGKRRKSNVSDDNKNPTSYRGVRMRSWGKWVSEI 86

Qy         53 ---RKKGCLRLGTHGSPTASSQSSATNMA 78
                 |||  : |||:  |||   : | ::|
Db         87 REPRKKSRIWLGTY--PTAEMAARAHDVA 113


RESULT 36
Q8VIM3
ID   Q8VIM3      PRELIMINARY;      PRT;   681 AA.
AC   Q8VIM3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Seven-span membrane protein FIRE.
GN   EMR4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=21448681; PubMed=11564768;
RA   Caminschi I., Lucas K.M., O'Keeffe M.A., Hochrein H., Laabi Y.,
RA   Kontgen F., Lew A.M., Shortman K., Wright M.D.;
RT   "Molecular cloning of F4/80-like-receptor, a seven-span membrane
RT   protein expressed differentially by dendritic cell and monocyte-
RT   macrophage subpopulations.";
RL   J. Immunol. 167:3570-3576(2001).
DR   EMBL; AF396935; AAL31879.1; -.
DR   MGD; MGI:1196464; Emr4.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   EGF-like domain.
SQ   SEQUENCE   681 AA;  76168 MW;  A833518D570CCD2C CRC64;

  Query Match             15.2%;  Score 64.5;  DB 11;  Length 681;
  Best Local Similarity   27.4%;  Pred. No. 82;
  Matches   23;  Conservative   15;  Mismatches   37;  Indels    9;  Gaps    3;

Qy          4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSV---AVEEGLAWRKKGCLRL 60
              ||     : |:  :||  |: :    |  : |::  | | |     ||| :|  |||  :
Db        251 SGAIRSEVKPV--LSEPVLLTL----QNIQPIDSRAEHLCVHWEGSEEGGSWSTKGCSHV 304

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
               |: | |       :: |:  : |
Db        305 YTNNSYTICKCFHLSSFAVLMALP 328


RESULT 37
Q91ZE5
ID   Q91ZE5      PRELIMINARY;      PRT;   689 AA.
AC   Q91ZE5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   EGF-like module-containing mucin-like receptor EMR4.
GN   EMR4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Stacey M.J., Chang G.W., Lin H.H.;
```

```
RT   "Mouse EMR4 a novel member of the EGF-TM7 family.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY032690; AAK51125.1; -.
DR   MGD; MGI:1196464; Emr4.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   EGF-like domain; Receptor.
SQ   SEQUENCE   689 AA;  77044 MW;  D9469A095CBC2088 CRC64;

  Query Match           15.2%;  Score 64.5;  DB 11;  Length 689;
  Best Local Similarity  27.4%;  Pred. No. 83;
  Matches   23;  Conservative   15;  Mismatches   37;  Indels    9;  Gaps    3;

Qy          4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSV---AVEEGLAWRKKGCLRL 60
              ||     : |:  :|| |: :    |  : |::  | | |     ||| :|  |||  :
Db        259 SGAIRSEVKPV--LSEPVLLTL----QNIQPIDSRAEHLCVHWEGSEEGGSWSTKGCSHV 312

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
               |: | |       :: |:  : |
Db        313 YTNNSYTICKCFHLSSFAVLMALP 336


RESULT 38
Q8BYX0
ID   Q8BYX0      PRELIMINARY;      PRT;   689 AA.
AC   Q8BYX0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical membrane all-alpha structure containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK037483; BAC29816.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   689 AA;  77084 MW;  88DE9A095CBC209B CRC64;
```

```
  Query Match             15.2%;  Score 64.5;  DB 11;  Length 689;
  Best Local Similarity   27.4%;  Pred. No. 83;
  Matches   23;  Conservative   15;  Mismatches   37;  Indels    9;  Gaps    3;

Qy          4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSV---AVEEGLAWRKKGCLRL 60
              ||     : |:  :||  |: :    |  : |::  | | |     ||| :|  |||  :
Db        259 SGAIRSEVKPV--LSEPVLLTL----QNIQPIDSRAEHLCVHWEGSEEGGSWSTKGCSHV 312

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
               |: | |       :: |:  : |
Db        313 YTNNSYTICKCFHLSSFAVLMALP 336


RESULT 39
Q9UBZ1
ID   Q9UBZ1      PRELIMINARY;      PRT;   733 AA.
AC   Q9UBZ1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   APC2 protein (Fragment).
GN   APC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=99147086; PubMed=10021369;
RA   van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA   Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT   "Identification of APC2, a homologue of the adenomatous polyposis coli
RT   tumour suppressor.";
RL   Curr. Biol. 9:105-108(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA   Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT   "Adenomatous Polyposis Coli Homologs in Mammals and Flies.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ012652; CAB61207.1; -.
DR   EMBL; AF128222; AAF01784.1; -.
DR   InterPro; IPR000225; Armadillo.
DR   Pfam; PF00514; Armadillo_seg; 7.
DR   SMART; SM00185; ARM; 5.
FT   NON_TER     733     733
SQ   SEQUENCE   733 AA;  80876 MW;  09E56BE5F7032BAD CRC64;

  Query Match             15.2%;  Score 64.5;  DB 4;  Length 733;
  Best Local Similarity   30.2%;  Pred. No. 89;
  Matches   19;  Conservative    6;  Mismatches   19;  Indels   19;  Gaps    2;

Qy         41 ALSVAVEEGLAWRKKGCL----------------RLGTHGSPTASSQSSATNMAIHR--- 81
              |:| : |  :| |: |||                 | |  |:| |       | |:|
Db        296 AMSSSPESCVAMRRSGCLPLLLQILHGTEAAAGGRAGAPGAPGAKDARMRANAALHNIVF 355
```

```
Qy          82 SQP 84
               |||
Db         356 SQP 358


RESULT 40
Q8TJS3
ID   Q8TJS3      PRELIMINARY;      PRT;  1004 AA.
AC   Q8TJS3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TPR-domain containing protein.
GN   MA3704.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE011082; AAM07059.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 19.
DR   SMART; SM00028; TPR; 18.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   1004 AA;  112398 MW;  51B5D3F7A777DD3D CRC64;

  Query Match           15.2%;  Score 64.5;  DB 17;  Length 1004;
  Best Local Similarity  29.7%;  Pred. No. 1.3e+02;
  Matches   22;  Conservative    9;  Mismatches   30;  Indels   13;  Gaps    3;

Qy          10 SISPMRSISENSLVAMDFS------GQKSRVIENPTEALSVAVEEGLAWRKKG--CLRLG 61
               || |     ||| :           |  || :|   :|| : :| | ||  ||    :||
Db         331 SIEP-----ENSCIMSGIGEIYYQLGDYSRALEAFEQALRLDIENGFAWNGKGNVLCKLG 385

Qy          62 THGSPTASSQSSAT 75
                :     : :|  |
Db         386 KYQEALEAYESLLT 399
```

Search completed: January 13, 2004, 16:22:14
Job time : 45.3307 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:17:58 ; Search time 13.8898 Seconds
(without alignments)
284.400 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	423	100.0	540	1	GRBE_HUMAN	Q14449 homo sapien
2	386	91.3	538	1	GRBE_RAT	O88900 rattus norv
3	383	90.5	538	1	GRBE_MOUSE	Q9jlm9 mus musculu
4	191	45.2	535	1	GRB7_MOUSE	Q03160 mus musculu
5	189	44.7	594	1	GRBA_HUMAN	Q13322 homo sapien
6	186	44.0	621	1	GRBA_MOUSE	Q60760 mus musculu
7	179	42.3	532	1	GRB7_HUMAN	Q14451 homo sapien
8	72.5	17.1	369	1	HEM3_PEA	Q43082 pisum sativ
9	69	16.3	235	1	GSPN_PSEAE	Q51575 pseudomonas
10	66.5	15.7	445	1	MDM2_BRARE	O42354 brachydanio
11	64.5	15.2	196	1	PAAY_ECOLI	P77181 escherichia
12	64.5	15.2	209	1	PYRE_COXBU	Q45918 coxiella bu
13	64	15.1	470	1	YJIR_ECOLI	P39389 escherichia
14	62	14.7	408	1	THIL_CANTR	P33291 candida tro
15	61.5	14.5	539	1	U7I5_MOUSE	Q925f4 mus musculu
16	61.5	14.5	2316	1	PTPZ_RAT	Q62656 rattus norv
17	61	14.4	589	1	C49A_DROME	Q9v513 drosophila

18	61	14.4	661	1	ATI2_VZVD	P09264 varicella-z
19	61	14.4	1317	1	GAP_CAEEL	P34288 caenorhabdi
20	60.5	14.3	389	1	SCWA_YEAST	Q04951 saccharomyc
21	60.5	14.3	396	1	VE2_HPV48	Q80923 human papil
22	60.5	14.3	401	1	VE2_HPV1A	P03118 human papil
23	60.5	14.3	462	1	LEU2_LISMO	Q8y5r7 listeria mo
24	60.5	14.3	614	1	NRD1_HUMAN	P20393 homo sapien
25	60.5	14.3	678	1	ABG1_HUMAN	P45844 homo sapien
26	60.5	14.3	886	1	SM6B_MOUSE	O54951 mus musculu
27	60.5	14.3	1541	1	ASX1_HUMAN	Q8ixj9 homo sapien
28	60.5	14.3	3038	1	TRIO_HUMAN	O75962 homo sapien
29	60	14.2	429	1	NOCT_MOUSE	O35710 mus musculu
30	60	14.2	977	1	DLP1_HUMAN	O14490 homo sapien
31	60	14.2	992	1	DLP1_RAT	P97836 rattus norv
32	59.5	14.1	1090	1	NIT4_NEUCR	P28349 neurospora
33	59	13.9	408	1	THIK_CANTR	P33290 candida tro
34	59	13.9	467	1	RXRG_CHICK	P28701 gallus gall
35	59	13.9	1067	1	BAB2_DROME	Q9w0k4 drosophila
36	59	13.9	1530	1	SCP2_HUMAN	Q9bx26 homo sapien
37	58.5	13.8	134	1	ACPS_BRUME	Q8yg72 brucella me
38	58.5	13.8	141	1	PSAD_GUITH	O78502 guillardia
39	58.5	13.8	382	1	HEM3_ARATH	Q43316 arabidopsis
40	58.5	13.8	573	1	ILVI_HAEIN	P45261 haemophilus
41	58.5	13.8	685	1	YGO4_YEAST	P53118 saccharomyc
42	58.5	13.8	779	1	CDC4_YEAST	P07834 saccharomyc
43	58	13.7	466	1	LEU2_BUCDN	O85072 buchnera ap
44	58	13.7	471	1	LEU2_BUCRP	P48573 buchnera ap
45	58	13.7	472	1	LEU2_BACSU	P80858 bacillus su

ALIGNMENTS

```
RESULT 1
GRBE_HUMAN
ID   GRBE_HUMAN      STANDARD;      PRT;   540 AA.
AC   Q14449;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Growth factor receptor-bound protein 14 (GRB14 adapter protein).
GN   GRB14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96218175; PubMed=8647858;
RA   Daly R.J., Sanderson G.M., Janes P.W., Sutherland R.L.;
RT   "Cloning and characterization of GRB14, a novel member of the GRB7
RT   gene family.";
RL   J. Biol. Chem. 271:12502-12510(1996).
CC   -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE
CC       AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE
CC       INTERACTION IS MEDIATED BY THE SH2 DOMAIN (BY SIMILARITY).
CC   -!- SUBUNIT: Binds to the ankyrin repeat region of TNKS2 via its N-
```

```
CC       terminus.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC       endosomes.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE LIVER, KIDNEY,
CC       PANCREAS, TESTIS, OVARY, HEART, AND SKELETAL MUSCLE.
CC   -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 1 Ras-associating domain.
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L76687; AAC15861.1; -.
DR   HSSP; P35235; 1AYA.
DR   Genew; HGNC:4565; GRB14.
DR   MIM; 601524; -.
DR   GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50200; RA; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain; Phosphorylation.
FT   DOMAIN      106    192       RAS-ASSOCIATING.
FT   DOMAIN      234    342       PH.
FT   DOMAIN      439    535       SH2.
SQ   SEQUENCE   540 AA;  60954 MW;  A8FCFC16D7437B47 CRC64;

  Query Match             100.0%;  Score 423;  DB 1;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-40;
  Matches   84;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 414

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              ||||||||||||||||||||||||
Db        415 GTHGSPTASSQSSATNMAIHRSQP 438
```

```
RESULT 2
GRBE_RAT
ID   GRBE_RAT       STANDARD;      PRT;   538 AA.
AC   O88900;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Growth factor receptor-bound protein 14 (GRB14 adapter protein).
GN   GRB14.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   MEDLINE=98421528; PubMed=9748281;
RA   Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., van Obberghen E.,
RA   Mauvais-Jarvis F., Girard J., Burnol A.-F.;
RT   "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT   actions.";
RL   J. Biol. Chem. 273:26026-26035(1998).
CC   -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE
CC       AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE
CC       INTERACTION IS MEDIATED BY THE SH2 DOMAIN.
CC   -!- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-
CC       terminus (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC       endosomes (By similarity).
CC   -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 1 Ras-associating domain.
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF076619; AAC61478.1; -.
DR   HSSP; P35235; 1AYA.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
```

```
DR   PROSITE; PS50200; RA; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain; Phosphorylation.
FT   DOMAIN      104    190       RAS-ASSOCIATING.
FT   DOMAIN      232    340       PH.
FT   DOMAIN      437    533       SH2.
SQ   SEQUENCE   538 AA;  60592 MW;  CEBC9037E7868EEF CRC64;

  Query Match             91.3%;  Score 386;  DB 1;  Length 538;
  Best Local Similarity   88.1%;  Pred. No. 3.4e-36;
  Matches   74;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              | || |||||:|||||:|||||||||||||:|||:|||||||||||||||||||||||
Db        353 QARSACSSQSVSPMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKKGCLRL 412

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              | |||||| ||||| |||:|||||
Db        413 GNHGSPTAPSQSSAVNMALHRSQP 436


RESULT 3
GRBE_MOUSE
ID   GRBE_MOUSE     STANDARD;      PRT;   538 AA.
AC   Q9JLM9; Q8VDI2; Q9CR03;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Growth factor receptor-bound protein 14 (GRB14 adapter protein).
GN   GRB14.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20179877; PubMed=10713090;
RA   Reilly J.F., Mickey G., Maher P.A.;
RT   "Association of fibroblast growth factor receptor 1 with the adaptor
RT   protein Grb14. Characterization of a new receptor binding partner.";
RL   J. Biol. Chem. 275:7771-7778(2000).
RN   [2]
RP   SEQUENCE OF 1-290 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic liver;
RX   PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
```

RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE OF 332-538 FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE
CC AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE
CC INTERACTION IS MEDIATED BY THE SH2 DOMAIN (By similarity).
CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-
CC terminus (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC endosomes (By similarity).
CC -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 Ras-associating domain.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.
CC --

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF155647; AAF43996.1; -.
DR   EMBL; AK010849; BAB27221.2; -.
DR   EMBL; AK010903; BAB27256.2; -.
DR   EMBL; BC021820; AAH21820.1; -.
DR   HSSP; P35235; 1AYA.
DR   MGD; MGI:1355324; Grb14.
DR   GO; GO:0005070; F:SH3/SH2 adaptor protein activity; IPI.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50200; RA; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain; Phosphorylation.
FT   DOMAIN      104    190       RAS-ASSOCIATING.
FT   DOMAIN      232    340       PH.
FT   DOMAIN      437    533       SH2.
SQ   SEQUENCE   538 AA;  60573 MW;  04ABD6CEB6ABC6CB CRC64;

  Query Match           90.5%;  Score 383;  DB 1;  Length 538;
  Best Local Similarity  86.9%;  Pred. No. 7.4e-36;
  Matches   73;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
              |||| |:|||:|||||:||||||||||||:|||||:|||||||||||||||||||||||
Db        353 QGRSACNSQSMSPMRSVSENSLVAMDFSGEKSRVIDNPTEALSVAVEEGLAWRKKGCLRL 412

Qy         61 GTHGSPTASSQSSATNMAIHRSQP 84
              | ||||:| ||||| |||:|||||
Db        413 GNHGSPSAPSQSSAVNMALHRSQP 436


RESULT 4
GRB7_MOUSE
ID   GRB7_MOUSE     STANDARD;      PRT;   535 AA.
AC   Q03160;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Growth factor receptor-bound protein 7 (GRB7 adapter protein)
DE   (Epidermal growth factor receptor GRB-7).
GN   GRB7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=93028373; PubMed=1409582;
RA   Margolis B., Silvennoinen O., Comoglio F., Roonprapunt C.,
RA   Skolnik E.Y., Ullrich A., Schlessinger J.;
RT   "High-efficiency expression/cloning of epidermal growth factor-
RT   receptor-binding proteins with Src homology 2 domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:8894-8898(1992).
CC   -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE EPIDERMAL
CC       GROWTH FACTOR RECEPTOR WHICH IS THEN INHIBITED. THE INTERACTION IS
CC       MEDIATED BY THE SH2 DOMAIN. ALSO BINDS TO ERBB2.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 1 Ras-associating domain.
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94450; AAA37733.1; -.
DR   PIR; C46243; C46243.
DR   HSSP; P35235; 1AYA.
DR   MGD; MGI:102683; Grb7.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000159; RA_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50200; RA; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain.
FT   DOMAIN       99    185       RAS-ASSOCIATING.
FT   DOMAIN      228    341       PH.
FT   DOMAIN      434    515       SH2.
SQ   SEQUENCE   535 AA;  59959 MW;  CD8C307864703645 CRC64;

  Query Match             45.2%;  Score 191;  DB 1;  Length 535;
  Best Local Similarity   59.7%;  Pred. No. 5e-14;
  Matches   43;  Conservative    8;  Mismatches   17;  Indels    4;  Gaps    2;

Qy         13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
              |:||:|:|:||||||||   |||:|| |||| |:||  |||||   ||     ||  | |
Db        366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKTNHRLSL---PTTCSGS 422
```

```
Qy          73 SATNMAIHRSQP 84
               |  : ||||:||
Db         423 S-LSAAIHRTQP 433


RESULT 5
GRBA_HUMAN
ID   GRBA_HUMAN     STANDARD;      PRT;   594 AA.
AC   Q13322; O00427; O00701; O75222; Q92606; Q92907; Q92948;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Growth factor receptor-bound protein 10 (GRB10 adaptor protein)
DE   (Insulin receptor binding protein GRB-IR).
GN   GRB10 OR GRBIR OR KIAA0207.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=96036069; PubMed=7479769;
RA   Liu F., Roth R.A.;
RT   "Grb-IR: a SH2-domain-containing protein that binds to the insulin
RT   receptor and inhibits its function.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:10287-10291(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Nantel A., Mohammad-Ali K., Sherk J., Posner B.I., Thomas D.Y.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RX   MEDLINE=99096036; PubMed=9881709;
RA   Angrist M., Bolk S., Bentley K., Nallasamy S., Halushka M.K.,
RA   Chakravarti A.;
RT   "Genomic structure of the gene for the SH2 and pleckstrin homology
RT   domain-containing protein GRB10 and evaluation of its role in
RT   Hirschsprung disease.";
RL   Oncogene 17:3065-3070(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction
```


OM protein - protein search, using sw model

Run on: January 13, 2004, 16:17:58 ; Search time 13.8898 Seconds